

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-91112

(43)公開日 平成9年(1997)4月4日

(51)Int.Cl <sup>6</sup>	類別記号	序内整種番号	P I	技術表示箇所
G 06 F	3/14	3 4 0	G 06 F	3/14
	3/02	3 7 0		3 4 0 A
G 10 L	3/00	5 7 1	G 10 L	3/00
				R
				5 7 1 H
				5 7 1 G

審査請求 実請求 請求項の数12 O L (全 30 頁)

(21)出願番号	特願平8-56813	(71)出願人	000003078 株式会社東芝 神奈川県川崎市幸区堀川町72番地
(22)出願日	平成8年(1996)3月14日	(72)発明者	河野 森之 大阪府大阪市北区大淀中1丁目1番30号 株式会社京芝関西支社内
(31)優先権主張番号	特願平7-205202	(72)発明者	池田 朋男 大阪府大阪市北区大淀中1丁目1番30号 株式会社京芝関西支社内
(32)優先日	平7(1995)7月19日	(72)発明者	知野 哲朗 大阪府大阪市北区大淀中1丁目1番30号 株式会社京芝関西支社内
(33)優先権主張国	日本 (JP)	(74)代理人	弁理士 外川 英明

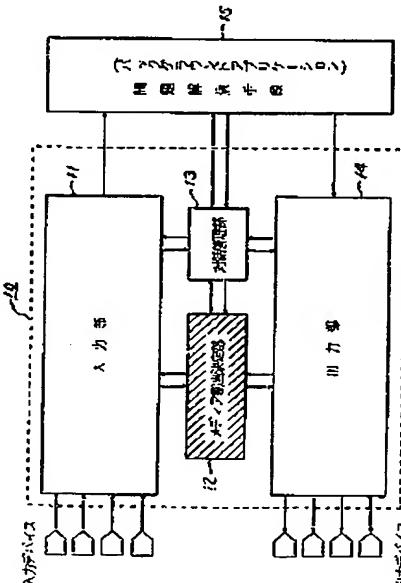
最終頁に続く

(54)【発明の名称】 マルチモードル対話装置及び対話方法

(57)【要約】

【課題】 入出力メディアを状況に応じて動的に切替えて対話する。

【解決手段】 異なる各種入力用デバイスのうち、入力メディア割当に対応して少なくとも一つのデバイスから利用者の入力データを解釈し、その解釈対応の入力を前記問題解決手段に与える入力手段11と、この入力対応に、システムの目的に応じた処理を行い、結果を出力する問題解決手段15と、この問題解決手段の出力を解釈し、この解釈結果を異なる各種出力用デバイスのうち、出力メディア割当に対応した出力用デバイスに振り分けて、利用者に输出する出力手段14と、利用者から複数の入力デバイスに与えられた入力データから問題解決手段への入力データを構築するために入力データを選択し解釈するための方法及びそれらのデータを組み合わせる方法を前記入力手段および出力手段による前記利用者との対話状況に対応して動的に生成するメディア割当手段12とを具備する。



1

## 【特許請求の範囲】

【請求項1】 利用者との対話により所要の処理を行って結果を出力する対話装置において、利用者からの情報を得るための異なる各種入力用デバイスのうち、入力メディア割当に従って少なくとも一つの入力用デバイスからの入力情報を解釈し、その解釈に基づいて対応情報を発生させる入力手段と、この発生された対応情報をそれぞれに、所要の目的に応じた処理を行い、結果を出力する問題解決手段と、前記問題解決手段の出力に基づいた結果を、異なる各種出力用デバイスのうち、出力メディア割当に従った出力用デバイスから出力する出力手段と、前記入力手段から前記問題解決手段へ与えられる前記対応情報を前記入力メディア毎に発生する前記対応情報を中から選択する手段、前記入力メディア毎に発生される対応情報を組み合わせる手段の少なくとも一つを割当てる前記入力メディア割当、および、前記問題解決手段から出力手段に出力する出力情報を得るために該問題解決手段からの情報を選択する手段およびそれらを組み合わせる手段の少なくとも一つを割当てる前記出力メディア割当、に基づいて前記利用者との対話状況に従い、入力または出力メディア割当を行うメディア割当手段と、を具備したことを特徴とするマルチモーダル対話装置。

【請求項2】 利用者との対話により所要の処理を行って結果を出力する対話装置において、利用者からの情報を得るための異なる各種入力用デバイスのうち、入出力メディア割当に従って少なくとも一つの入力用デバイスからの入力情報を解釈し、その解釈に基づいて対応情報を発生させる入力手段と、この発生された対応情報をそれぞれに、所要の目的に応じた処理を行い、結果を出力する問題解決手段と、前記問題解決手段の出力に基づいた結果を、異なる各種出力用デバイスのうち、入出力メディア割当に従った出力用デバイスから出力する出力手段と、前記入力手段から前記問題解決手段へ与えられる前記対応情報を中から選択する手段、前記入力メディア毎に発生される対応情報を組み合わせる手段の少なくとも一つ、および、前記問題解決手段から出力手段に出力する出力情報を得るために該問題解決手段からの情報を選択する手段およびそれらを組み合わせる手段の少なくとも一つ、の双方を対応させて割り当てる入出力メディア割当に基づいて前記利用者との対話状況に従い、入出力メディアの割当を行うメディア割当手段と、を具備したことを特徴とするマルチモーダル対話装置。

【請求項3】 利用者との対話により所要の処理を行って結果を出力する対話装置において、利用者からの情報を得るための異なる各種入力用デバイスのうち、入出力メディア割当に従って少なくとも一つの入力用デバイスからの入力情報を解釈し、その解釈に

2

に基づいて対応情報を発生させる入力手段と、この発生された対応情報をそれぞれに、所要の目的に応じた処理を行い、結果を出力する問題解決手段と、前記問題解決手段の出力に基づいた結果を、異なる各種出力用デバイスのうち、入出力メディア割当に従った出力用デバイスから出力する出力手段と、前記問題解決手段からの処理結果に応じて入出力メディア割当の変更処理を指示し、また、前記入力手段と出力手段による入出力を管理する対話管理手段と、10 利用者から少なくとも一つの入力デバイスに与えられた入力データから前記問題解決手段への入力を得るために、所定の規則に従って前記入力用デバイスのうち、利用するデバイスと入力データの入力方法の組合せの少なくとも一つ、および、前記問題解決手段の出力から利用者に指示する表現を得るために、所定の規則に従って前記出力用デバイスのうち、利用するデバイスと出力データの出力方法の組合せの少なくとも一つ、の双方を対応させて選択し割り当てるための入出力メディア割当を実施すると共に、入出力メディア割当の変更処理の指示を受けた時はその変更を行なうべく処理して新たなメディア割当をするメディア割当手段と、を具備したことを特徴とするマルチモーダル対話装置。

【請求項4】 入力対応に、システムの目的に応じた計算処理を行ない、結果を出力する問題解決手段と、異なる各種入力用デバイスのうち、入出力メディア割当に対応して少なくとも一つのデバイスから利用者の入力データを解釈し、その解釈対応の入力を前記問題解決手段に与える入力手段と、前記問題解決手段からの処理結果を得て、これを異なる各種出力用デバイスのうち、入出力メディア割当に対応した出力用デバイスに振り分けて、利用者に出力する出力手段と、前記問題解決手段からの処理結果に応じて入出力メディア割当の変更処理を指示し、また、前記入力手段と出力手段による入出力を管理する対話管理手段と、前記問題解決手段に与えられる入力情報を、入力メディア毎に発生する前記対応情報の中から選択する手段、および入力メディア毎に発生される対応情報を組み合わせる手段の少なくとも一つ、および前記問題解決手段の出力から利用者に指示する表現を生成するためにその出力を選択する手段およびそれらを組み合わせる手段の少なくとも一つ、の双方の組み合わせである前記入出力メディア割当を選択すると共に、入出力メディア割当の変更処理の指示を受けた時はその変更を行なうべく処理して新たな入出力メディア割当を選択するメディア割当手段と、を備えることを特徴とするマルチモーダル対話装置。

【請求項5】 前記対話管理手段には、入力装置および出力手段の少なくとも一方の動作状況を適宜監視し、現

3

在選択され利用されている入出力メディア割当の利用に対する障害が発生した場合に、該障害の発生を表す障害情報を、メディア割当決定手段に通知するメディア監視機能を備えると共に、  
前記メディア割当手段には該障害情報を受けると、入出力メディア割当を、適宜変更する機能を備えることを特徴とする請求項2または3または4いずれか1項記載のマルチモーダル対話装置。

【請求項6】 前記問題解決手段には、システムの問題解決のために必要な情報を利用者から獲得する必要があるような場合に、前記対話管理手段に対し入力要求情報を発生し、前記入力要求情報を利用者に入力してもらうよう求める入力要求機能。

もしくは、前記対話管理手段には、前記出力手段が設定されている入出力メディア割当に従って利用者に対する出力を構成するために必要な情報を、問題解決手段に出力を求める出力要求情報を発生し、前記出力要求情報を問題解決手段が outputするよう求める出力要求機能。のうち少なくとも一つの機能を備えると共に、前記メディア割当手段には、入出力メディア割当変更処理の指示を受けた時は、前記入力要求情報を、もしくは前記出力要求情報をうち少なくとも一つを利用して新たなメディア割当を行なう機能を備えることを特徴とする請求項3または4いずれか1項記載のマルチモーダル対話装置。

【請求項7】 利用者との対話により利用者から得た情報を解釈して入力対応情報を発生させ、この発生された入力対応情報を所要の目的に応じた処理を行ない結果を出力する問題解決装置に通信手段を通じて入力対応情報を送りし、問題解決装置から通信手段を通じて得られた該入力対応情報を対する処理結果を利用者との対話により出力する対話装置において。

利用者からの情報を得るための異なる各種入力用デバイスのうち、入出力メディア割当に従って少なくとも一つの入力用デバイスからの入力情報を解釈し、その解釈に基づいて対応情報を発生させる入力手段と、

通信手段を通じて前記問題解決装置から得られた結果を、異なる各種出力用デバイスのうち、入出力メディア割当に従った出力用デバイスから出力する出力手段と、前記問題解決装置に対し、入力手段において発生された前記入力対応の送付、および、問題解決装置が発生し出力した問題解決結果の受けとりを行なう通信手段と、前記入力手段から前記問題解決装置へ送付される前記対応情報を前記入力メディア毎に発生する前記対応情報の中から選択する手段、前記入力メディア毎に発生される対応情報を組み合わせる手段の少なくとも一つ。

および、前記問題解決装置から出力手段に出力する出力情報を得るために該問題解決装置からの情報を選択する手段及びそれらを組み合わせる手段の少なくとも一つ、の双方の組合せである入出力メディア割当を前記利用者との対話状況に従い発生し、利用者との対話を行なうた

10

4

めの入出力メディア割当を行なうメディア割当手段と。

を具備したことを特徴とするマルチモーダル対話装置。

【請求項8】 利用者との対話により利用者から得た情報を解釈して入力対応情報を発生させ、この発生された入力対応情報を所要の目的に応じた処理を行ない結果を出力する問題解決装置に入力対応情報を送付し、問題解決装置から得られた該入力対応情報をに対する処理結果を利用者との対話により出力する対話装置において。

19

利用者からの情報を得るための異なる各種入力用デバイスのうち、入出力メディア割当に従って少なくとも一つの入力用デバイスからの入力情報を解釈し、その解釈に基づいて対応情報を発生させる入力手段と、

通信手段を通じて前記問題解決装置から得られた結果を、異なる各種出力用デバイスのうち、入出力メディア割当に従った出力用デバイスから出力する出力手段と、前記問題解決装置からの処理結果に応じて入力メディア割当、出力メディア割当の変更処理を指示し、また、前記入力手段と出力手段による入出力を管理する対話管理手段と、

20

前記問題解決装置に対し、入力手段において発生された前記入力対応の送付、および、問題解決装置が発生し出力した問題解決結果の受け取りを行なう通信手段と、利用者から少なくとも一つの入力デバイスに与えられた入力データから前記問題解決装置に与える入力対応情報を得るために、所定の規則に従って前記入力用デバイスのうち、利用するデバイスと入力データの入力方法の組合せの少なくとも一つ。

39

および、前記問題解決装置から得られた結果から利用者に提示する表示を得るために、所定の規則に従って前記出力用デバイスのうち、利用するデバイスと出力データの出力方法の組合せの少なくとも一つ、の双方を対応させて選択し割り当てるための入出力メディア割当を実施すると共に、入出力メディア割当の変更処理指示を受けた時はその変更を行なうべく処理して新たなメディア割当をするメディア割当手段と、

40

を具備したことを特徴とするマルチモーダル対話装置。

【請求項9】 前記メディア割当手段には入出力メディア割当を記録するメディア割当記憶スロットを持ち、前記メディア割当記憶スロットに現在選択されている入出力メディア割当を記録すると共に、

入出力メディア割当変更処理の指示を受けた時は、前記メディア割当記憶スロットに記録されている入出力メディア割当情報を利用して新たなメディア割当を行なう機能を備えることを特徴とする請求項2または3または4または7または8いずれか1項記載のマルチモーダル対話装置。

50

【請求項10】 前記メディア割当手段には、過去に選択された入出力メディア割当の系列を記憶するメディア

割当ヒストリ記憶スロットを持ち、前記メディア割当ヒストリ記憶スロットにそれまでに選択されてきた入出力メディア割当の系列を記憶し、入出力メディア割当が変更された時は、そのときに選択されていた入出力メディア割当を前記メディア割当ヒストリ記憶スロットに追加すると共に、  
入出力メディア割当変更処理の指示を受けた時は、前記メディア割当ヒストリ記憶スロットに記録されている過去の入出力メディア割当情報をを利用して新たなメディア割当を行なう機能を備えることを特徴とする請求項2または3または4または7または8いずれか1項記載のマルチモーダル対話装置。

【請求項11】前記対話管理手段には、前記問題解決装置がシステムの問題解決にのために必要な情報を利用者から獲得する必要があるような場合に、問題解決装置が発生し、対話管理手段に対して通信手段を通じて送付された入力要求情報をに対し、前記入力必要情報を利用者に入力してもらうよう前記入力手段に求める入力要求受け付け機能、

もしくは、対話管理手段には、前記出力手段が設定されている入出力メディア割当に従って利用者に対する出力を構成するために必要な情報を、問題解決装置に出力を求める出力要求情報を発生して前記通信手段を通じて送付し、前記出力必要情報を問題解決装置が通信手段を通じて送付してくるよう求める出力要求機能、  
のうち少なくとも一つの機能を備えると共に、

前記メディア割当手段には、入出力メディア割当変更処理の指示を受けた時は、前記入力要求情報を、もしくは前記出力要求情報をうち少なくとも一つを利用して新たなメディア割当を行なう機能を備えることを特徴とする請求項8記載のマルチモーダル対話装置。

【請求項12】利用者との対話により所用の処理を行って結果を出力する対話方法において、

利用者からの情報を得るために異なる各種入力用デバイスのうち、入出力メディア割当に従って少なくとも一つの入力用デバイスからの入力情報を解釈し、その解釈対応の情報を発生させる入力ステップと、

この発生した情報対応に、所要の目的に応じた処理を行い、結果を出力する問題解決ステップと、

前記問題解決ステップとの出力に基づいた結果を、異なる各種出力用デバイスのうち、入出力メディア割当に対応した出力用デバイスから出力する出力ステップと、

前記問題解決ステップからの処理結果に応じて入出力メディア割当の変更処理を指示し、また、前記入力ステップと出力ステップによる入出力を管理する対話管理ステップと、

前記複数の入力用デバイスのうち、少なくとも一つの入力デバイスに与えられた入力データから前記問題解決ステップへの入力を得るために、前記入力用デバイスのうち、利用するデバイスを選択し、入力メディア毎に発生

される入力の組合せの少なくとも一つ、  
および、前記問題解決手段の出力から利用者に提示する表現を生成するために前記出力用デバイスのうち、利用するデバイスと出力データの出力方法の組合せの少なくとも一つ、  
の双方の組合せである前記入出力メディア割当を所定の規則に従って選択し、該入出力メディア割当を実施すると共に、これら入出力メディア割当の変更処理の指示を受けた時はその変更を行なへて処理して新たなメディア割当を生成するメディア割当ステップと、  
を備えることを特徴とするマルチモーダル対話装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、情報処理装置において、組合せて利用可能な複数の入出力手段を有するヒューマン・マシン対話装置およびその制御方法に関する。

【0002】

【従来の技術】近年、電子計算機は、キーボード、マウス、マイクロフォン、タッチパネル、イメージスキャナ、ビデオカメラ、ペン、データグローブ、対センサなどの種類の異なる複数の入力手段、すなわち入力デバイスを装備し、様々な情報を入力することができるようになっており、また、ディスプレイやスピーカといった出力デバイスにも、様々な情報を音声言語や効果音、音楽といった様々な形態で出力できる機能を装備してきており、これらの各種入力デバイスを有効に利用したり、使い易いヒューマン・マシンインタフェイスの要求が高まっている。また、近年、これらの入力デバイスを用いて、例えば利用者がディスプレイ上の対象を指で指示しながら音声でコマンドを発声するなど、複数の入力デバイスを組み合わせた複雑な入力が可能な入力手段、すなわち入力メディアを装備し、また、ディスプレイやスピーカなどの複数の出力デバイスを利用して、例えばスピーカから音声言語でメッセージを伝えながらディスプレイに表示した人物の顔の表情と、スピーカからの効果音でニュアンスを伝えるといった具合に、複数の出力デバイスやその出力内容および形式を組み合わせた複雑な出力が可能な出力手段、すなわち出力メディアを装備することで、インタフェイスの使い易さや情報伝達の質及び効率を向上させようというマルチモーダルインターフェイス技術の開発が盛んに行われている。

【0003】ところで従来、スムーズで自然な、利用者と計算機上のアプリケーションソフトウェアとの間の情報の授受を実現するためには、この情報の授受を利用者とアプリケーションの間の対話と見なし、予め記録された対話規則に基づいて対話プランを展開し、対話規則において決められた入出力方法の組合せでマルチモーダル対話を実現するなどしていた。

【0004】

【発明が解決しようとする課題】しかし、各々の対話のフェイズにおいて入出力に使用するデバイスとデバイスの利用方式の組合せ、すなわち、メディア割当（メディアアロケーション）を対話規則中に予め記述しておく方式では、例えば、周辺ノイズ音疊が大きく音声を入出力手段として用いることが適切でないような状況に対処できないなど、メディアアロケーションが対話の流れに固定的であり、状況に応じて入出力手段の組合せを選択し利用者と対話することができないという問題があった。

【0005】例えば、利用者が音声で入力することが可能な対話装置を利用する場合を考えてみる。この場合、現在の技術では音声認識の技術が十分成熟していないため、利用者によっては何度も同じ単語を発生しても認識に失敗するということが発生し得る。これは、発音に癖があり過ぎる場合等に多く発生し、対話装置側での認識成功の確率は極めて低くなる。

【0006】にも拘らず、従来のマルチモーダル対話装置はこのような利用者に対しても、音声による入力の直しを何度も要求してしまうため、時間の無駄が生じたり、利用者にストレスを感じさせたりするという問題があった。

【0007】また、例えば音声による地名入力に対して音声認識により、該当の観光地をデータベースから検索し、検索結果を出力手段に出力するといった構成の音声指示による旅行案内システムを考えてみると、この旅行案内システムが利用者に音声による地名入力を求めたが、これに対して利用者が例えば「竜安寺」（りょうあんじ）という地名を「りゅうあんじ」と読み間違えて発声したため、システムが検索に失敗したような場合、従来のマルチモーダル対話装置においては入出力メディア割当を動的に行う手段を持たないため、利用者とのコミュニケーションを回復するのが困難であった。

【0008】つまり、利用者が「りょうあんじ」という正しい有名な地名の読みを知らないことが、利用者とのコミュニケーションの失敗に繋がっているわけであるが、これも例えば、漢字で入力選択する形式に切り替えたりすれば、コミュニケーション回復が出来る可能性が極めて大であるにもかかわらず、このような入出力メディア割当を状況に応じて変更するといったことができないので、硬直状態に陥り、結果的には案内サービスができないことになる。

【0009】また、例えば従来からある対話装置においてデータベース検索結果の内容を利用者に提示するような場合、出力形態は予めプログラミングされた形態に固定されている。つまり、検索結果の出力件数にかかわりなく同一の出力形態が保られ、例えば数十件あるいは数百件の検索結果があるのに、それを一つ一つ音声で読み上げたり、逆に数件しかしない検索結果を表形式で表示したりといった、利用者にとってわかりにくく使いにくい入出力方式となる場合が少なからずあった。

【0010】さらに、利用者が要求するメディア割当に合せてシステムが利用者と対話を行うことができないため、システムの利用者はシステムが決定し指定したメディア割当に従って入出力を行わなければならないという問題もあった。

【0011】入力手段には種々のものがあるが、利用者にとって使い易い入力手段は各人共通しているわけではなく、例えば、ある利用者は音声による入力を好みが、別の利用者はコマンド文字列をキーボード入力するのを好みといった具合に、利用者により使い易い入出力手段は異なっているのが普通である。

【0012】しかしながら、従来の対話装置においてはこれらの個々人の差異が考慮されておらず、固定的な入出力手段の組合せが提供されていて、利用者の好みで入出力手段を選択的に利用できるようになっておらず、使い勝手が良いものとはいえないという問題があった。

【0013】また、入力手段を複数持つ従来のマルチモーダル対話装置において、どの入力手段を用いるべきかといったことや、どの入力手段がある時点で利用可能かというこを利用者が認識することは困難であり、入力に際して利用者に戸惑いやストレスを与えるという問題があった。

【0014】更に、例えば従来の音声による入力もしくは出力が可能な対話装置において、外部の騒音が激しくなるなど音声を利用するのか好ましくない場合でも音声を入力あるいは出力に用いるなど、動的にメディアアロケーション（メディア割当）が行なえない従来の対話装置においては、外部環境的要素の変化に応じた入出力方式の変更が行えないという問題があった。

【0015】このように従来のマルチモーダル対話装置では、利用者はシステムが持つ対話規則で予め決められた入出力モードの組合せに従ってシステムと対話を行うことを強制させ、利用者にとってわかり易く使い易いインターフェイスではなかったため、利用者が情報の入力に際してストレスを感じたり、出力を理解し難かったりしており、ときには入出力の失敗の原因ともなっていた。

【0016】また、音声単語認識モードでは認識失敗に起因する入力や失敗や誤入力があり得るといったように、選択された特定の入出力手段の性質に起因する入出力の失敗の場合でも、結果として対話の目的が未達成となってしまっていた。

【0017】そこで、この発明の目的とするところは、利用者の使い易い入出力インターフェイスとなり得るマルチモーダル対話装置を提供することにある。特に本発明では、ある時点で適切な入出力のメディア割当を動的に選択する機能を持つことで、アプリケーションプログラムのフロントエンドとして様々な状況に応じて利用者に使い易い入出力インターフェイスを提供することができるマルチモーダル対話装置を作成することにある。

【課題を解決するための手段】本発明に係るマルチモーダル対話装置は、利用者から複数の入力デバイスに与えられた入力データからアプリケーションプログラムへの入力データを精緻するために入力データを選択し解釈するための方法及びそれらのデータを組み合わせる方法を予め定めたルールに基づいて動的に生成する手段と、アプリケーションプログラムの出力データから利用者に提示する表現を生成するために出力データを選択する方法及びそれらのデータの組み合わせ・表現の方法を予め定めたルールに基づいて動的に生成する手段を具備し、利用者との対話の各々の時点において、入出力のそれぞれについて適切な手段の組合せを動的に選択し対話を維続することができることを特徴とするものである。

【0019】本発明によれば、利用者が計算機システムと対話をを行う際に、対話の各時点で利用可能な入出力手段を反映した入出力方式を選択し、また利用者がその利用者にとって利用しやすい入出力手段を用いた入出力の方法を選択し利用することが可能となる。

【0020】また本発明によれば、メディアアロケーション（メディア割当）が対話の流れに固定的であり、状況に応じて入出力手段の組合せを選択し利用者と対話することができないという従来のマルチモーダル対話装置の問題を解決することが可能となる。

【0021】更に本発明によれば、従来のマルチモーダル対話装置ならば特定の利用者に対して音声による入力のし直しを何度も要求してしまうなど、時間の無駄が生じたり利用者にストレスを感じさせたりするという問題があったような場合においても、動的に入出力手段を選択し切替える手段を具備することによりこの問題を解決し、利用者が時間の無駄やストレスを感じることなく計算機システムを利用することが可能となる。

【0022】更に本発明によれば、入力の解釈処理が失敗した場合に従来のマルチモーダル対話装置においては利用者とのコミュニケーションを回復するのが困難であったのに対し、動的に入出力手段を選択し切替える手段を具備することにより、この問題を解決し、速かに利用者とのコミュニケーションを回復することができる。

【0023】更に、例えば従来からある対話装置においてデータベース検索結果の内容を利用者に提示するような場合、少ない検索結果を知らせるために音声出力の形態を採用していれば、検索結果が数十件、数百件あってもその検索結果の出力を音声で読み上げるなど、出力件数にかかわりなく同一の出力形態が採られ、逆に大盛の検索結果の扱いを容易にするために表形式を採用している場合には、数件しかない検索結果であっても表形式で表示したりといった固定的なものである。

【0024】しかし、本発明では動的に入出力手段を選択し切替える手段を設けて、このような、利用者にとってわかりにくく使いにくい入出力方式となるような場合においては、動的に入出力手段を選択し切替えるように

することにより、この問題を解決し、利用者にわかりやすい入出力方式をとることができるようになる。

【0025】更に、本発明によれば、従来の対話装置においては使い易い入出力手段に関する個々の差異が考慮されておらず、固定的な入出力手段の組合せが選択されていたという問題に対して、各々の利用者がそれぞれにとって使いやすい入出力方法の組合せを利用することが可能となる。

【0026】更に、本発明によれば、入力手段を複数持つ従来のマルチモーダル対話装置において、どの入力手段を用いるべきかやどの入力手段がある時点で利用可能かということを利用者が認識することは困難であり、入力に際して利用者に戸惑いやストレスを与えるという問題に対して、利用者が入力に利用することが可能な手段を表示もしくは掲示する手段を具備することによりこの問題を解決し、利用者が入力手段を容易に認識することが可能となり、利用者にわかりやすく使いやすいインターフェイスを提供することが可能となる。

【0027】かくして、利用者は利用者とアプリケーションプログラムが求める情報の質と量、及び利用する計算機システムが利用可能な入出力手段に依存するがアプリケーションプログラムに依存しないヒューマン・マシンインターフェイスを利用することが可能になる。

【発明の実施の形態】本発明は、入出力のそれについて、手段の組み合わせを動的に選択する手段を具備することにより、状況に応じた入出力方法を利用者に提供することで、使い易いインターフェイスを実現し、また、利用可能な入力手段を利用者に知らしめる手段を具備することで、よりわかり易い入力環境を利用者に提供し、さらに入出力手段の組み合わせを変更して利用者との対話を失敗から回復し、対話を維続することができるマルチモーダル対話装置を提供するものであり、以下、図面を参照して本発明の詳細について説明する。

【0029】図1は本発明の具体例に係るマルチモーダル対話装置の構成図である。図1に示すように本発明のマルチモーダル対話装置10は入力部11、メディア割当決定部12、対話管理部13、出力部14の計4つの構成要素からなり、システムに接続される物理デバイスを通して利用者との対話処理を行い、利用者と問題解決手段（バックグラウンドアプリケーション）15の仲立ちを行う。

【0030】入力部11はマウスデバイス、マイクロフォン、テレビカメラ、キーボード、タブレット、イメージキャナ、センサなど各種の入力デバイス $ODV_a \sim ODV_n$ の出力を取り込むためのものであり、出力部14はディスプレイ、スピーカ、ランプ、プリンタ、プロッタ等各種の出力デバイス $ODV_a \sim ODV_n$ への駆動出力を発生するものである。

【0031】メディア割当決定部12は入力用出力用の

11

各種デバイスとそれら各種デバイスの利用方式の組合せ、すなわち、メディア割当（メディアアロケーション）を決定するための装置であり、その時々での最良の組み合わせを選択することができる機能を有する。

【0032】また、対話管理部13は問題解決手段15との間で接受される情報や、入力部11および出力部14からの状態情報等に基づいてメディア割当の要求を行ったり、対話のプランを生成したりする等、ユーザとシステムの間の対話の方法、流れを制御するものである。

【0033】問題解決手段15は、入力に対してその回答や返答を行なうアプリケーションであり、これは対話によって目的の回答を出したり、目的の仕事をするようなアプリケーションであって例えば、観光案内システムや定期券発券機等なんでも良い。図15に、本発明のマルチモーダル対話装置10の詳細構成をブロック図で示す。

【0034】（入力部11および出力部14の詳細）図15に示すように、本発明のマルチモーダル対話装置10の入力部11は入力デバイス処理部111、入力モード処理部112、入力メディア処理部113の3つのモジュールから構成され、また、出力部14は出力メディア処理部141、出力モード処理部142、出力デバイス処理部143の3個のモジュールからなる。

【0035】ここで、モードとは一つのデバイスからの入力（出力も同様）を解釈し、抽象化したものとする。例えば、マウスデバイスから入力されたデータはそれを解釈するモード部品により操作対象の指示、コマンドの発行といった複数のモードによる解釈と利用が可能である。

【0036】同様に、メディアとは複数（一つでもよい）のモードからの解釈結果を組み合わせたものとし、入力を組み合わせて得られる処理結果を指示内容として指示対象に渡すことができるものを指している。

【0037】例えば、マウス・オブジェクト指示-音声コマンドメディアは、パソコン等のポインティングデバイスであるマウスデバイスからの入力と、音声によるコマンドを併せて解釈してコマンド発行をするものであり、アプリケーションへのコマンドを指示する音声を解釈することのできる音声コマンドモードと、マウスデバイスからの入力をコマンド操作対象を指示する操作として解釈するマウス・オブジェクト指示モードの解釈結果とを入力とし、それらの入力を組み合わせて「マウスで指示された対象に対して音声で指示されたコマンドを発行する」という処理をアプリケーションに伝えるといったことを行うメディアである。

【0038】<入力部11>入力部11の構成要素である入力デバイス処理部111は、各物理入力デバイスから伝達されたデータを入力とし、デジタルデータへの変換等そのデータを解釈するための最小限の処理を行う入力デバイス処理部品を構成要素に持つ。デバイス処理

12

部品は入力デバイス毎に用意され、独立・並列に動作する。各デバイス処理部品は、対応するデバイスへの入力に上記の処理を施した後、対応する一つまたは複数の入力モード処理部112に送付する。例えばマイクデバイス処理部品は、例えば利用者がマイクロフォンから入力した音声を背景ノイズと共に取り込み、例えば12kHzの標準周波数、16bitの量子化ビット数のデジタルオーディオ信号に変換し、その信号を入力モード処理部112に出力する。

【0039】入力モード処理部112は、各抽象入力モードへの入力を処理する部品を構成要素に持つ。一つの入力モード処理部品は一つの抽象入力モードに対応し、一つの入力デバイス処理部から送付された整形されたデータを、抽象モード毎に抽象化し解釈する。各入力モード部品は入力デバイス部品と例えば表1に示したように1対1、もしくは1対多の対応関係にある。

【0040】各入力モード処理部112で解釈された結果は、入力メディア処理部113に送付される。例えば、音声コマンド入力モード処理部品は、送られてきたデジタルオーディオ信号を、例えば高速フーリエ変換による周波数分析により特徴パラメータ時系列を求め、さらに例えば複合類似度法、隠れマルコフモデルなどの方程式に基づき音声認識を行い、与えられた認識対象コマンド語彙の中から尤度の最も高い語彙を得、その情報を入力メディア処理部113に出力する。

【0041】入力メディア処理部113は、入力モード処理部112から得られた利用者の入力のモード毎の解釈結果を各抽象メディア単位に組合せることで統合された入力を生成し、得られた複数のメディア入力のうちから必要なメディアからの入力をメディア割当決定部12の指示に従いながら選択し、問題解決手段15に対する入力コマンドおよびデータを与える。

【0042】入力メディア処理部113の構成要素は、各抽象メディア毎に用意された入力メディア処理部品113aと、それらの部品の処理結果のうちどれを問題解決手段15に送付するかを指定するメディア選択スイッチSWである。各入力メディア処理部品113aは、一つまたは複数の入力モード部品からの出力データ（抽象モードでの解釈結果）を入力として受け入れ、それらを統合してメディアとしての解釈結果をメディア選択スイッチSWに送る。

【0043】メディア選択スイッチSWは、対話管理部13の指示に従いバックグラウンドアプリケーションに送付するメディア出力を選択し、送付する。例えばマウス・オブジェクト指示-音声コマンドメディア入力処理部品は、マウス・オブジェクト指示モード入力処理部品の出力と音声コマンドモード入力処理部品の出力を入力とし、それらを組み合わせてメディアとしての解釈を生成し出力する。そして、例えば、現在の入力メディアとしてマウス・オブジェクト指示-音声コマンドメディア

50

13

が選択されている場合、メディア選択スイッチSWはこのメディアによる入力の解釈結果のみを問題解決手段15に引き渡す。

【りり44】<出力部14>出力部14の構成要素である出力メディア処理部143は、対話管理部13が与えるメディアのアロケーション情報(割当情報)に従いながら問題解決手段15からの出力データを選択し、各抽象出力メディアに振り分けてメディア毎に加工し、それらを出力モード単位に分割・統合して出力モード処理部の各モード処理部品にその出力モードが取り扱うデータを受け渡す。このような処理を行つために、出力メディア処理部143は出力メディア振り分けスイッチ143a、出力メディア処理部143b、メディア・モードインターフェイス143cの3つのモジュールから成る。メディア振り分けスイッチ143aは、バックグラウンドアプリケーションが送付してきた出力データを対話管理部13の指示に従いながら、実際に(利用者に)出力されるデータを選択し、それを抽象メディア処理部品に振り分ける。

【りり45】出力メディア処理部品143は、各抽象出力メディアに対応する出力メディア処理部品を構成要素として持ち、メディア振り分けスイッチ143aにより担当メディアに振り分けられた出力データを、そのメディアに出力できる形に加工して送出する。メディア・モードインターフェイス143cは、各抽象出力モード1に対応するメディア・モードインターフェイス部品を構成要素に持つ。各メディア・モードインターフェイス部品はそのモードが担当する1つまたは複数の抽象メディア処理部品からの出力を統合し、対応する出力モード処理部品にデータを受け渡す。

【りり46】出力モード処理部142は、各抽象出力モードへの出力を処理する出力モード処理部品142aとモード毎の出力データを出力デバイス毎に統合するモード・デバイスインターフェイス142bを構成要素に持つ。一つの出力モード処理部品は一つの抽象出力モードに対応し、そのモードに対応するメディア・モードインターフェイス部品から送られてきたメディアデータに出力モード毎の処理を施し、モード・デバイスインターフェイス142bに受け渡す。

【りり47】モード・デバイスインターフェイス142bの一つの部品は一つの出力デバイス部品に対応し、そのデバイスが担当する1つまたは複数の抽象デバイス処理部品からの出力を統合し、対応する出力デバイス処理部品にデータを受け渡す。出力デバイス処理部141は、各出力デバイス毎に用意された出力デバイス処理部品を構成要素として持つ。各出力デバイス処理部品は、対応するメディア・デバイスインターフェイスの出力データを入力コマンド及びデータとして担当する出力デバイスを駆動する。

【りり48】(メディア割当決定部12の詳細)図15

10

14

に示されるように本発明のマルチモーダル対話装置10のメディア割当決定部12は、符号121を付して示したように、メディア割当生成部といふモジュールと、符号122を付して示すメディア割当選択テーブル、および符号123を付して示す現メディア割当記憶スロット、符号124を付して示すモードメディア割当対応表の3つの記憶部品から構成される。これらのうち、メディア割当生成部121は、実際にアプリケーションに入力するデータ及びその組み合わせ(メディアアロケーション)、そして対話管理部13の問い合わせに応じて次メディア割当選択テーブル112を参照して、実際にユーザに与えるデータとしその適切な組み合わせ・表現方法を生成し、対話管理部13に伝達する。

20

【りり49】(対話管理部13の詳細)図15に示されるように、本発明のマルチモーダル対話装置10の対話管理部13は、入力メディア処理部113および出力メディア処理部143から得られる入出力モードの状態と、問題解決手段(バックグラウンドアプリケーション)15とのデータ交換の状態を、問題解決手段15から問題解決手段15の状態の通知や問題解決手段15からの入出力要求を、複数の利用者に共通した特徴情報、問題解決手段15の制御方法に関する情報をイベントとして受け取り、それらの情報をもとに対話のプランを生成する。

29

【りり50】そして、そのプランの実行過程において入力メディア処理部113と出力メディア処理部143、そして問題解決手段15に指示を行うことにより、ユーザとシステムの間の対話の方法・流れを制御する。

39

【りり51】このような処理を行う構成としては、入力メディア処理部113や出力メディア処理部143、問題解決手段(バックグラウンドアプリケーション)15、およびメディア割当決定部12といった他のモジュールとの間の情報の授受を対話と見做せば、例えば、特開平5-216618号に開示した音声対話システムのように、状態遷移モデルを用いた方法や、Sacerdoti, E. D.著“*A Structure for Plans and Behavior*”, American Elsevier刊, 1977年にあるような階層プラン合成方式を利用することで、実現可能である。

49

【りり52】対話管理部13は符号131を付して示したように、対話プランナと呼ぶモジュールを持ち、この対話プランナ131は対話規制記憶テーブル132、および対話プリミティブ記憶テーブル133の2つの記憶テーブルと、図示しないがプラン記憶部、初期目標記憶スロット、および現在目標スロットという3つの記憶部品を持ち、それにより対話プランを構築し、実行する。

59

【りり53】以上が本装置の構成とその機能である。本装置の中核モジュールである対話管理部13の動作のフローチャートを図19に示す。この図に基づいて本装置

の動作原理とデータの利用方法について述べる。

【0054】まず、本装置が起動されると、対話管理部13の対話プランケ131はプラン生成のための初期目標が予め設定されている初期目標記録スロットから取り出し、その目標を基に対話管理部13内の図示しないプラン記憶部に全ての対話プランを順に展開する。

【0055】プラン記憶部の一つの項目は、“ノードID”、“目標”、“状態”、“サブノードIDリスト”、“結果”の5つのスロットから成り、これをプランノードと呼ぶ。ある一つの目標が立てられると、プランノードが一つ生成され、設定された目標が目標スロットに代入され、ノードIDスロットには各プランノードにユニークな記号が割り当てられる。

【0056】状態スロットには、“未処理”、“処理中”、“成功”、“失敗”、“再試行”の5つのうち1つが代入される。サブノードIDリストスロットには、目標スロットの内容から展開されたサブ目標に対応するプランノードのIDがリストの形式で収納される。

【0057】結果スロットには、そのスロットが展するプランノードの目標の実行結果が収納される。以降、例えば、“ノードIDスロットの値がp000であるプランノード”のことを、単に“プランp000”、あるいは“サブプランp000”と記述する。

【0058】ある対話目標をそのサブ目標に展開するために、対話管理部13は対話規則記憶テーブル132を持つ。対話規則記憶テーブル132には図2に示すような対話規則を表す複数の項目が収められており、その各々を対話規則ノードと呼ぶ。一つの対話規則ノードは“ルールID”、“目標”、“適用条件”、“サブ目標リスト”の4つのスロットからなる。“ルールID”はプランノードのノードIDと同様に、各対話規則にユニークな記号が割り当てられる。

【0059】各対話規則ノードには、その規則が達成する目標が目標スロットに納められ、目標が展開されて生成される（一般に複数の）サブ目標の系列がサブ目標リストスロットに納められ、規則を適用するための条件が適用条件スロットにそれぞれ納められる。

【0060】対話プランケ131は対話規則記憶テーブル132に収められた対話規則を上から順に探索し、条件に適合する最初の対話規則を適用する。対話プランの生成と実行は、与えられた目標を満足する規則を対話規則の中から探し、適用してサブ目標に展開・実行するという操作を再帰的に繰り返すことで実現されている。

【0061】そして、最終的に初期目標が満足された場合、対話全体が成功し、本装置は停止する。以降、例えば、ルールIDスロットの値が“r00”である対話規則ノード”のことを、単に“対話規則r00”、あるいは“ルールr00”と記述する。

【0062】対話プランケ131はある時点での対話の目標を現在目標スロットに納めて管理する。現在目標ス

ロットの目標を展開する対話規則を探索する前に、対話プランケ131は図3に示すような直接実行可能な対話の最小単位である対話プリミティブを収納した対話プリミティブ記憶テーブル133を探査する。

【0063】対話プリミティブ記憶テーブル133の一つの項目を対話プリミティブノードと呼び、これは“ファクトID”、“目標”、“仕様”的3つのスロットからなる。現在目標スロットの内容が、対話プリミティブ記憶テーブル133中の、ある対話プリミティブノードの目標スロットの値に单一化可能である場合、対話プランケ131はその目標を対話プリミティブとなみし、対応する対話プリミティブノードの仕様スロットに記述されている内容を実行する。

【0064】対話管理部13が管理し実行するモジュール間の対話のプリミティブは、メディア割当決定部12に対する適切なメディア割り当ての問い合わせとその回答、入出力メディア部品に対する設定指令と入出力指令、問題解決手段15に対する入出力要求のいずれかである。

【0065】以降、例えば、「ファクトIDスロットの値が“f00”である対話プリミティブノード」のことを単に“対話プリミティブf00”あるいは“ファクトf00”と記述する。

【0066】対話プランの展開・実行時にある時点での目標に单一化可能な対話プリミティブ、および対話規則のいずれもなかった場合、その目標は失敗する。ある目標が失敗した場合、対話プランケ131はコンピュータプログラム言語のひとつである“Prolog言語”的実行過程において計算機をバックトラックするのと同様の方法で、対話プランの展開過程の後戻りと対話規則の再試行を行う。

【0067】すなわち、ある対話目標P<sub>1</sub>の実行に際し、対話規則R<sub>1</sub>を適用して失敗した場合、対話プランケ131 P<sub>1</sub>に適用可能な他の対話規則、もしくは対話プリミティブを探査し、適用可能なものがあればその対話規則または対話プリミティブを適用してP<sub>1</sub>を満足するという目標を再試行する。

【0068】P<sub>1</sub>に適用可能なすべての対話規則および対話プリミティブが失敗した場合、対話プランケ131は対話目標P<sub>1</sub>の直前に成功した対話目標P<sub>2</sub>まで後戻りし、同様にその対話目標P<sub>2</sub>の再試行を行う。目標達成の失敗が続く限り目標の後戻りは連続的に発生し、仮に初期目標まで後戻りして初期目標 자체が失敗した場合、対話全体が失敗したことになり、本装置は停止する。

【0069】メディア割当決定部12は、マルチメディア割当(MA)の問い合わせである“MA問合せ”という対話プリミティブの実行という形で対話プランケ131から呼び出される。メディア割当決定部12は、現メディア割当記憶スロットに収納されている問い合わせが発生し

17

た時点でのメディア割当の状態と、図6に示すような次メディア割当選択テーブル122を用いて、その時点で適切なメディア割当を決定する。

【0070】次メディア割当選択テーブル122の一つの項目を次メディア割当選択ノードと呼び、各次メディア割当選択ノードは“ノードID”、“現メディア割当”、“条件”、“次メディア割当”的4つのスロットを持つ。

【0071】メディアアロケーション生成部121は、現メディア割当記憶スロットの内容とメディア割当決定部12が呼び出された際の条件が適合する次メディア割当選択ノードを、次メディア割当選択テーブル122の中から探し、次メディア割当の候補を生成する。

【0072】次にメディア割当生成部121は、発見した次メディア割当選択ノードの次メディア割当スロットの内容から、図7に示すようなモード～メディア割当対応表124を探索して、あるメディア割当を適用する際の制約条件を検証する。モード～メディア割当対応表124の一つの項目をモード～メディア割当対応ノードと呼び、各モード～メディア割当対応ノードは、“ID”、“メディア割当名”、“制約”、“入出力モード組”的4つのスロットからなる。

【0073】メディア割当生成部121がメディア割当の制約検証を行う際には、メディア割当名のスロットが探索するメディア割当名と合致するモード～メディア割当対応ノードを探し、そのモード～メディア割当対応ノードの制約スロットに記述されているメディア割当を実際に用いる際の制約を取り出し、問い合わせ時点での状態がその制約を満足しているかどうかを検証する。

【0074】この制約が満足されていれば、そのメディア割当候補が最終的に次メディア割当として選択され、対話プランナ131に回答される。ここで、メディア割当生成部121がメディア割当の決定を依頼され、次メディア割当選択テーブルを探索する際に、あるメディア割当ノードの現メディア割当スロットもしくは条件スロットの値が“-”の場合、そのスロットの値は検証に用いられない。

【0075】例えば、現メディア割当スロットの内容が“-”である次メディア割当ノードは、現メディア割当記憶スロットの値が何であっても条件スロットの内容とモード～メディア対応表の制約スロットの内容が適合していれば、適応可能である。また、次メディア割当ノードの現メディア割当スロットおよび次メディア割当スロットには、メディア割当の内容以外に入力メディアと出力メディアの組の形でも記述できる。

【0076】また、この入力メディアと出力メディアのいずれかに、変数を用いることが可能である。例えば、図6にノードIDスロットの値が“m06”である次メディア割当選択ノードにおいて、現メディア割当スロット値が“（音声言語、XXX）”であり、次メディア割

19

29

39

49

59

18

当スロットの値が“（ベン文字列、XXX）”となっているが、次メディア割当選択ノードは現メディア割当記憶スロットの値になっているメディア割当の入力メディア、すなわち、現在設定されている入力メディアが“音声言語”であれば、出力メディアが何であれ他の制約が満足されれば適用可能である。

【0077】このとき、現メディア割当スロットの出力メディア側の値、すなわち、変数“XXX”には現メディア割当記憶スロットの出力メディア側の値が一時的に代入され、その代入結果はその現メディア割当スロットが属する次メディア割当選択ノードに反映される。

【0078】そして、次メディア割当スロットの制約の探索と検証は、モード～メディア割当対応表の入出力メディアスロットの値と対応づけることにより行う。例えば、現メディア割当スロットの値が“ヒューマン”である時に、呼出条件が“入力失敗”でメディア割当生成部121が呼び出されたとする。

【0079】メディア割当生成部121は次メディア割当を決定するために、まず次メディア割当選択テーブル122を探査するが、図6にある次メディア割当選択ノードのうちノードIDスロットの値が“m01”から“m05”的次メディア割当選択ノードは条件が適合しないため選択されない。

【0080】次にメディア割当生成部121はノードIDスロットの値が“m06”的次メディア割当選択ノードが適応可能であるか否か検証しようとするが、この次メディア割当選択ノードの現メディア割当スロットの値が入力メディアと出力メディアの組の形式で記述されているため、まず現メディア割当記憶スロットの値を入力メディアと出力メディアの組に再構成する。

【0081】そのために、メディア割当生成部121は図7を探査し、ノードIDスロットの値が“m01”的モード～メディア割当対応ノードのメディア割当スロットの値が現メディア割当記憶スロットの値と同じ“ヒューマン”であることから、このモード～メディア割当対応ノードの入出力モードスロットの値からメディア割当“ヒューマン”的入力メディアが“音声言語”モード単体であり出力メディアが“音声言語”モードと“顔画像”モードの組合せである“（音声言語、音声言語+顔画像）”という入出力メディア組であることを得る。

【0082】メディア割当生成部121はこの情報を用い、ノードIDの値がm06である件の次メディア割当選択ノードの現メディア割当スロットの値である、

“（音声言語、XXX）”と单一化を図り、変数XXXに“音声言語+顔画像”を一時的に代入する。

【0083】この代入は同じ次メディア割当選択ノードの次メディア割当スロットに反映され、このスロットの値は“（ベン文字列、音声言語+顔画像）”となる。次に、メディア割当生成部121は同じ次メディア割当選

択ノードの条件スロットの値「呼出条件=入力失敗」を検証し、この場合は成功する。さらにメディア割当生成部121は図7の如きモード-メディア割当対応表の深さを行い、入出力モードスロットの値が“(ベン文字列、音声言語+顔画像)”であるモード-メディア割当対応ノードの制約を検証する。

【0084】仮にその制約が満足された場合、次メディア割当の選択は成功し、このモード-メディア割当対応のノードのメディア割当名スロットの値である“ベン文字列入力・ヒューマン対応”と言うメディア割当が現メディア割当記憶スロットに代入され、メディアアロケーション結果（メディア割当結果）として対話管理部13に返される。変数XXXXへの一時的な代入は、その変数が所要する次メディア割当選択ノードの検証が終了する・と解消される。以降、例えば「ノードIDスロットの値が“m00”である次メディア割当選択ノード」のことを、単に“次メディア割当選択ノードm00”と記述する。また例えば「IDスロットの値が“mm00”であるモード-メディア割当対応ノード」のことを、単に“モード-メディア割当対応ノードmm00”と記述する。

【0085】【具体的動作例（第一の動作例）】本発明を理解し易くするために、具体例として本装置を観光地における旅館案内のアプリケーションに適用した場合の例をつぎに述べる。ここで観光地における旅館案内のアプリケーションが、すなわち、今までの説明で登場した問題解決手段15に相当する。

【0086】対話管理部13の対話規則記憶テーブル132には、図2に示す対話規則および図3に示す対話プリミティブが予め記憶されている。また、初期目標記憶スロットには、“対話セッション”という初期目標が予め与えられ記憶されている。

【0087】まず本装置を起動させた時、対話プランナ131は初期目標を初期目標記憶スロットから取り出し、その目標を満足する規則“r00”を呼び出す。そして、対話プランナ131によって対話規則“r00”は図4に示すように展開される。

【0088】対話プランナ131は初期のサブ目標であるユーザ検出を呼び出し、対話規則“r11”が展開され、さらに対話規則“r12”が展開される。サブ目標“MA問合せ（待機、MA1）”の実行にあたり、対話プランナ131は適応可能な対話規則が対話規則記憶テーブル132に無いため、図3に示す対話プリミティブ記憶テーブル133を検索する。

【0089】その結果、対話プランナ131は、対話プリミティブ“f01”を発見し、仕様の項目にある“メディア割当決定部呼び出し”を実行し、メディア割当決定部12にその時点で呼び出し条件に適合し、かつ、利用可能なメディア割り当てを問い合わせる。

【0090】メディア割当決定部12は、その時点のメ

ディア割当と利用可能なメディアやその他の状況をもとに、図6に示すような次メディア割当選択テーブル122を引いて次に選択すべきメディア割当を決定する。この場合テーブル要素“m01”が適用され、“待機”的メディア割当が選択される。

【0091】次にメディア割当決定部12は、図7のモード-メディア割当対応表を引いて該当するメディア割当を適用する際の制約を確認する。この場合、入力センサデバイスが利用可能であれば、メディア割当決定部12は最終的に“待機”的メディア割当を選択し対話管理部に返答する。

【0092】対話管理部13は、メディア割当決定部12が返答してきたメディア割当に基づき、入力メディア処理部113および出力メディア処理部143に、それを利用すべきメディアを通知する。これにより、入力メディア処理部113および出力メディア処理部143はその通知に従ってモード-メディア割当対応表を引き、それぞれメディア選択スイッチSW、メディア振り分けスイッチ143aの状態を変更することで入出力チャネルを設定する。

【0093】この例の待機メディア割当の場合、入力チャネルとして対入センサのみを開き、出力チャネルは例えば図20のような待機メッセージを画面に出力する。以上の処理により、サブ目標“p11”は成功し、対話プランナ131は次のサブ目標“検出待機”を実行に移そうとする。このサブ目標は図3の対話プリミティブ記憶テーブル133に記憶されている通り対話プリミティブであり、即座に実行される。

【0094】このとき、対話管理部13内のプラン記憶部の状態は図5のようになっている。対話プリミティブ“ユーザ検出”は、対入センサ検出メディアの状態が変化するまで待ち状態を続ける。

【0095】該箇の前に利用者が立つと入力センサが反応し、対話プリミティブ“ユーザ検出”が成功し、サブプラン“p01”が満足され、次のサブ目標“MAR選択（初期）”が呼び出される。このサブ目標は更に“r12”により展開され、待機メディアの選択時と同様に対話プランナ131はサブ目標“MA問合せ”を実行するためにメディア割当決定部12を呼び出して対話開始のためのメディア割当を問合せせる。

【0096】メディア割当決定部12は、待機メディア割当の場合と同様に、自己の保持する次メディア割当選択テーブル122とモード-メディア割当対応表124を引くことによって、条件と制約を満足するメディア割当を選択する。

【0097】この結果、この例の場合には“ヒューマン”というメディア割当が選択され、対話管理部13に返答される。対話管理部13は返答されたメディア割当を入力メディア処理部113と出力メディア処理部143に指示することで、メディア割当“ヒューマン”に基

づく対話を行う。

【0098】このメディア割当“ヒューマン”というの は、入力に音声認識、出力には顔画像と音声言語を用いることで音声による自然な対話が行えるメディア割当である。従って、利用者は自分の戸で質問や要望、回答や応答等を行えば良く、それに対する返答や回答等は装置側がディスプレイに顔画像を表示し、併せて音声言語を用いて行ってくれることになって、より人間がわかり易く、扱い易いものとなる。

【0099】また戸に、例えば上のような場合において、背景雑音が大きいために音声入力モードもしくは音声出力モードが使用不可能であると入力メディア処理部113もしくは出力メディア処理部143からメディア割当決定部12にデータが与えられていれば、メディア割当“ヒューマン”は選択されず、別のメディア割当が行われて対話が継続される。例えば、背景雑音の影響を全く受けることのないように入力はキーボード、出力は画面による文字表示といった具合である。

【0100】かくして、本発明は従来のマルチモーダル対話装置の基本構成に、適切なメディア割当を決定する手段を追加する構成としたことにより、対話の状況およびその時点で利用可能な入出力手段に合わせて適切な入出力メディアを効率的に選択することができるようになり、これにより、状況に応じてより使い易いインターフェイスを利用者に提供することが可能となる。

【0101】以上は、マルチモーダル対話装置の側がその時々の状況に対応して最適な種々の入出力メディアを効率的に選択することで、使い易いインターフェイスを利用者に提供するものであったが、利用者の側が種々の入出力メディアのうちの、使用したいものを任意に選択指定することで、個々の利用者の使い易いインターフェイスを提供する構成とすることもできる。

【0102】マルチモーダル対話装置において、利用者の側が種々の入出力メディアのうちの、使用したいものを任意に選択指定する構成にするには、前述の構成に更に、利用者が入力可能なメディアを掲示する手段を付加すれば良い。このような構成とすることにより、利用者がどのような手段で入力を行うことができるかを容易に知ることができ、より使い易いインターフェイスを提供することが可能となる。

【0103】例えば、上記の動作例のメディア割当“ヒューマン”においては、図16のように例えばマイクのアイコンを画面に表示することで音声による入力が可能であることを利用者に知らしめ、音声による入力を促す効果が期待できる。その後、サブ目標“開始セッション”が対話規則“r03”で満足される過程で、利用者のセッション開始要求とシステムの認識が行われ、かくして図8に示す一例としてのシナリオの“A)”および“B)”を動作させることが可能となる。

【0104】すなわち、シナリオの“A)”

案内システムの前に一人の旅行者が立つ

これにより、案内システムではセンサがこれを検知してディスプレイをオンにし、システムのタイトル画面を表示し、効果音と共に入力を促すプロンプトを画面に表示する。

【0105】シナリオの“B)”

旅行者が音声により「宿を探して下さい」と話しかけ、これにシステムが「了解しました」と答える。

【0106】このような内容の開始セッションが成功すると、対話管理部13は問題解決手段である案内システムを実現するアプリケーションにタスクが開始されたことを通告する。この場合、問題解決手段である当該アプリケーションは、その通知に応じてデータベース検索のための初期検索条件の獲得、すなわち、“到着日”、“出発日”、“入数”、“予算”を対話管理部13に対して要求する。

【0107】この要求に基づいて、対話管理部13は再びメディア割当決定部12に適切なメディア割当の決定を要請し、その解答に基づき、メディア割当“案内表示音声入出力”をセットし、図17に示すような画面のように入力中の検索条件の状態を表示しながら、図8のシナリオの“C)”に示す如き対話を音声を用いて行うことで、対話管理部13は問題解決手段であるアプリケーションの要求を満足する。

【0108】(第二の動作例) 次に場面展開を変えた第二の動作例について述べる。図8のシナリオ“C)”までの対話により、データベース検索の初期条件を得て、問題解決手段(バックグラウンドアプリケーション)は検索処理を行い、検索終了後、対話管理部13に対して検索結果の出力を要求する。

【0109】このとき、問題解決手段15側は出力要求と共に検索結果の概要、すなわち検索結果の件数と1件毎のデータベース要素、すなわち、“名称”、“名称の読み”、“住所”、“規模”、“部屋の種類と各部屋の宿泊料”、その他“特徴”、“外観写真”、“地図”、“そこまでの行き方と到着時間”、“最寄り駅”、“付近の名勝”が問題解決手段15側からマルチモーダル対話装置10の出力部14における出力メディア処理部143に出力されることが通告される。

【0110】このとき、対話プランナ131のプラン記憶部は図10のサブ目標“p42”まで成功した状態となっている。次のサブ目標“p43”である“候補決定”を満足するために、対話プランナ131は対話規則を検索し、まず“r07”を適用しようとするが適用条件が合わないために失敗し、他の対話規則を検索する。

【0111】次に対話プランナ131は“r08”を選択し、サブ目標“p431”である“選択要求”が展開される。このサブ目標は同様に対話規則“r10”によって、サブ目標系列[p4311, p4312, p4313]に展開される。

【0112】これらのサブ目標系統列のうち、サブ目標“p4311”は更に展開されてプラン記憶部は図10のような状態になり、更に対話規則“r12”に従い、メディア割当決定部12が呼び出される。メディア割当決定部12ではこれまでの記述と同様の処理が行われるが、検索結果が3件と少ないことから次メディア割当選択テーブル122の深さにおいて次メディア割当選択ノード“m3”が条件に適合して選択される。

【0113】その結果、利用者が検索結果の中から候補を選択するための入出力メディア割当として“タッチパネル表示選択・要素名音声読み上げ”が選択され、対話管理部13により、そのメディア割当が実行される。

【0114】これにより、図18のように画面上に3件の宿泊施設についてそれぞれ“名称”、“住所”、“写真”、“規模”、“付近の名称”が配置されたタッチパネルボタンを、出力メディア処理部143が問題解決手段15から流れてきたデータから選択してセットするとともに、図8のシナリオ“D)”のように、それぞれの物件の名称を音声で読み上げてその選択を促すメッセージがスピーカから音声が入出力されることになる。

【0115】従来からある対話装置の場合、出力件数にかかわらず同一の出力形態が取られ、例えば数十件あるいは数百件の検索結果を音声で読み上げたり、逆に数件しかない検索結果を表形式で表示したりといった、利用者にとってわかりにくく使いにくい入出力方式となる場合が少なからずあった。

【0116】例えば、データベースの検索結果が多いような場合にも、従来の対話装置では図8のシナリオ“D)”の対話と同様に全ての物件の名称が読み上げられるといったことが発生していた。

【0117】本対話装置はこのような問題にも対処すべく、従来からある階層プラン生成方式や状態遷移グラフ方式に基づく対話管理機構に加えて、状況に応じ動的に入出力の形態を設定する機能を導入している。従って、本装置によれば、対話管理部13がメディア割当決定部12にその時点で適切なメディア割当を問い合わせ、その回答内容に対応してメディア割当を動的に設定し、利用者に提示する物件数が多ければ、例えば、表形式の検索結果提示を選択するといったように、その時々の状況に合わせて最適な形態で検索結果提示をすることができます。

【0118】例えば、図8のシナリオ“D)”の対話においては、上述のように検索結果の提示物件数が3件であり、この3件の宿泊施設の検索結果から利用者に並みのものを選択させるために、メディア割当生成部121で次メディア割当選択テーブル122のノード“m04”が選択され、その結果、図18のように1件に対して比較的画面上に並んで表示される形での入出力メディア割当がなされていた。

【0119】しかしながら、同様の場面でも検索結果件

数が仮に10件であった場合、メディア割当生成部121で選択される次メディア割当選択ノードは“m05”となり、これによって図21のような表形式の画面出力が割り当てられ、また宿泊施設名を音声で読み上げることなく検索結果件数のみを音声で伝えるといった形で利用者にストレスを与えないような入出力メディア割当がなされる。

【0120】かくして、本装置によればアプリケーション（問題解決手段）が出力を要求する情報の質と量に対応して利用者に受け入れ易く、わかり易い出力手段の組み合わせを設定することが可能となる。

【0121】更に、例えば図8のシナリオ“G)”の場合のように、データベース検索結果の件数が極めて多いような場合には、表形式での出力形態でも利用者が結果を選択する際に大きな負担を与えるため、アプリケーション（問題解決手段）の要求に従って検索結果を提示すること自体が適切でない。

【0122】この場合にはつぎのようになる。今、上記図8のシナリオ“D)”のケースと同様に、問題解決手段による検索処理が行われ、対話プランナ131のプラン記憶部は図10のようにサブ目標“p43111”まで対話が進んでいるとする。

【0123】この例の場合、検索結果要素数が534件と多く図6の次メディア割当選択テーブルに条件が適合する次メディア割当選択ノード122がないため、メディア割当決定部12を呼び出すサブ目標“p43111”、すなわち“MA問い合わせ”が失敗してしまう。“p43111”的失敗を受けて、対話プランナ131は周知のコンピュータ言語の一つであるp�o1o8インタプリタと同様の戻り値処理を行う。

【0124】すなわち、“p43111”的上位ゴールである“p4311”的“MA選択（選択肢提示）”にまず戻りし、“p4311”に適用可能な対話規則もしくは対話プリミティブがないか再試行する。この場合、失敗した対話規則“r12”意外に適用可能な対話規則も対話プリミティブもないため、“p4311”も失敗し、“p4311”的上位ゴールである“p431”的“選択要求”まだ戻りする。

【0125】更にこの“p431”も同様に失敗するため、“p43”的“候補決定”まで戻りし、同様にこのゴールの再試行のために適用可能な対話規則もしくは対話プリミティブが選択され、図2の対話規則“r09”が適用され、図11のようにサブ目標系列[p4301, p43012, p43013]が展開される。

【0126】対話プランナ131は“p4301”的“検索条件絞り込み”的実行に際し、更に、対話規則“r14”を適用し、サブ目標系列[p43011, p43012, p43013]を生成する。

【0127】次に対話プリミティブであるサブ目標“p43011”的“未入力変数問合せ”的実行過程で対話

25

管理部13は問題解決手段15に対し、物件検索にまだ用いていない変数を問合せ、問題解決手段15側はその要請に対し、未入力の変数リスト【観光予定地、部屋様式、宿泊施設様式】を報告する。

【0128】次のサブ目標“p43012”的“mem  
ber”(入力変数、【観光予定地、部屋様式、宿泊施設様式】)の実行で未入力変数リストの内から変数“観光予定地”が取り出される。

【0129】そして、その変数を追加条件として利用者に入力させるために、次のサブ目標“p43013”的“追加条件入力”対話規則“r17”が適用され、サブ目標系列【p430131, p430132, p430133】が生成される。

【0130】これらのサブ目標系列の実行過程において図8のシナリオ“G)”のような対話を行われ、かくして本装置はメディア割当の失敗から回復して利用者との対話を継続することが可能となる。

【0131】図11は、図11図8のシナリオ“G)”のシステムの発話S14が発せられた時のプラン記憶部の状態を示している。

【0132】(第三の動作例)次に場面展開をえた第三の動作例について述べる。図8のシナリオ“D)”までの対話の後に、図8のシナリオ“E)”のように利用者が道順と到達時間を求めたとする。すなわち、対話装置11が決定し出力したメディア割当に対して利用者がそのメディア割当とは異なるメディア割当での出方を求めるような場合、従来のマルチモーダル対話装置においては入出力メディア割当を効的に行う手段を持たないため、そもそもこのような利用者の要求に応えることが不可能であった。しかし、本発明の装置ではこのような事態に対応できる。

【0133】このような事態に遭遇した場合、本装置は対話の制御と入出力メディアの割り当てを分離し、メディア割当決定部12が対話の流れとは独立にメディア割当選択することができる。そのため、このような事態での対処を可能にしている。

【0134】図8のシナリオ“E)”におけるU6のような利用者の要求があった時、対話プランナ131のプラン記憶部の状態は図12のよう、サブ目標“p4313”的“ユーザ選択”の処理中の状態で利用者の入力を待っている状態であった。ここで利用者の入力が候補の選択好意でなく、メディア割当変更とそれに伴う再出力要求であったため、サブ目標“p4313”は失敗する。

【0135】この失敗に伴い第二の動作例の場合と同様に後戻り処理が発生し、“p4313”的親目標である“p431”的“選択要求”が失敗し、図2の対話規則“r15”が再試行される。この対話規則の実行過程で対話プランナ131は“ユーザ要求(出力([道順、到達時間]))”という条件で、メディア割当決定部12

26

に適切なメディア割当の選択を依頼する。この依頼により、メディア割当決定部12は“名称、写真、地図タッチパネル表示選択・道順音声読み上げ”というメディア割当を選択する。そして、この選択したメディア割当を対話プランナ131に渡す。

【0136】対話プランナ131はこの選択に従い、宿泊施設の選択肢を表示する際の出力メディア割当を変更し、三件の各宿泊施設についてその“名称”、“写真”、“地図”をディスプレイに表示してタッチパネルボタンを設定すると共に、図8のシナリオ“E)”におけるS11のような発話【駿山山荘はバスで約1時間半ほどです。ひなびた庵には列車で30分ほど行ったところからバスに乗り換えて30分程かかります。奥山院はここからバスで30分ほど乗ってから30分ほど歩いたところにあります。】といった内容の発話をを行い、利用者に情報を提供し、宿泊施設の候補の選択を再び利用者に要求する。この時点でのプラン記憶部の状態を図13に示す。

【0137】かくして、本装置によれば利用者が要求する入出力方法を用い、また利用者の割込みに応じて利用者が主導権を持った形での対話継続が可能となり、より個々の利用者に使い易いインターフェイスを提供することが可能となる。

【0138】(第四の動作例)利用者から得た音声入力の内容が、正しくないためにシステムが音声認識処理に失敗したような場合での対処動作例を、次に、第四の動作例として述べる。図8のシナリオ“H)”のように、システムが利用者からの音声入力を求めたが、利用者が地名を読み間違えたため、システムが音声認識処理に失敗したような場合、従来のマルチモーダル対話装置においては入出力メディア割当を効的に行う手段を持たないため、利用者とのコミュニケーションを回復するのが困難であった。

【0139】しかしながら、このような場合にも本装置では、対話プランナ131が利用者との対話の失敗を対話規則実行時の失敗と捉え、失敗したサブ目標を満足できる他の対話規則を探索し、適用可能な対話規則があれば再試行する。

【0140】この例では音声入力というヒューマン・マシンインターフェイスとしての観点がはぐれていますが、入力処理、特に読み間違いが発生し易い地名の入力処理の確実性という観点では劣っている入力メディアにおいて失敗が発生したことから、手書き文字認識という入力メディアを選択し再試行することで利用者との対話を継続することに成功している。

【0141】このような制御は次のようにして実現されている。既に述べたように、図8のシナリオG)に示したシステム発話S14つまり、「どのようなところを観光されるご予定でしょうか?」という発話の時点でのプラン記憶部の状態は、図11のようになっている。

【0142】ここで、音声言語モードでの入力処理において、利用者が“竜安寺”を指名するつもりで“りゅうあんじ”と読み間違えたとする。この“りゅうあんじ”は音声認識の対象語彙となっていたなかったとすると、音声認識処理が失敗してしまうことになる。

【0143】このため、サブ目標“p 4301331”は失敗し、その親目標“p 430133”的“入力処理”は適用可能な別の対話規則“r 19”で再試行される。この“r 19”は同じメディア割当で再入力を求めるという対話規則であり、この対話規則により、“p 4301331”は図4に示すようにサブ目標系列【p 43013301, p 43013302】に展開され、それらの実行過程で装置と利用者のあいだでS 15とS 9の対話がなされる。

【0144】つまり、  
装置：“は？もう一度お願いします”  
利用者：“りゅうあんじです”  
なる対話がなされる。

【0145】しかし同様の理由で音声認識処理は失敗し、サブ目標“p 4301331”は再び失敗する。次に適用された規則“r 20”は、入力失敗をキーとしてメディア割当をやり直し、再入力させる対話規則であり、この対話規則により“p 4301331”は再試行され、サブ目標列【p 43013311, p 43013312, p 43013313】に展開される。

【0146】サブ目標“p 43013311”的MA選択（入力失敗（観光予定地））は更に対話規則“r 12”でサブ目標系列【p 430133111, p 430133112】に展開される。

【0147】サブ目標“p 430133111”的“MA問合せ”は対話プリミティブであり、対話プリミティブ記憶テーブル133の内容に従いメディア割当決定部12の呼び出しが行われる。

【0148】既に述べたように、メディア割当決定部12は図6に示す次メディア割当選択テーブル122を参照し、次メディア割当選択ノード“m 6”がこの場合選択される。この次メディア割当選択ノードは現在のメディア割当の入力モードのみを音声言語モードからペン文字列入力モードに変更するというメディア割当の選択であり、このメディア割当に従って変換入力の対話を継続することで、図8のS 16以降に対話を進め、更には宿泊施設の選択というタスクを継続させるような対話に復帰することが可能となる。

【0149】すなわち、  
装置：“認識できませんでした。ペンで行き先の名前を書いてください”  
利用者がペンで画面に「竜安寺」と書くことで文字認識し、装置は「竜安寺」を「りょうあんじ」と認識する。  
装置：“竜安寺（りょうあんじ）ですね。検索しますのでしばらくお待ちください”

なる対話が成功する。

【0150】かくして、本装置によれば利用者との対話の失敗に対して、別の入出力手段に替えて再度実施すべく、入出力手段を変更するよう制御する構成としたことで失敗から回復し、利用者との対話を継続することが可能となる。

【0151】更に、例えば外部の騒音が激しくなったような場合に対処できるように、外部騒音を捉えるデバイスを本装置に付加し、その入力により音声入出力の適切性を判断する入力モード部品を実現すれば、メディア割当生成部121でメディア割当を行う際に、音声入力可能もしくは音声出力可能と図7のモード～メディア対応表の割約部に記述されているメディア割当については、選択されることがなくなる。

【0152】その結果、例えば音声入力をを行うメディア割当が選択されていて、利用者からの入力を音声で受け付けていた時に急に騒音が激しくなったような場合に、上記の動作例と同様に音声認識失敗に伴ってメディア割当決定部呼び出しが発生し、メディア割当決定部での次メディア割当の選択が新たに実施されるが、その選択においては音声を入出力に用いるようなメディア割当は選択されることがなくなる。

【0153】かくして、このような構成の装置によれば、外部環境的要素の変化に応じて入出力方式を変更する制御を付加することで、外部環境的要素に応じて最適な入出力方式を選択して使用するので、利用者との対話を継続することが可能となる。

【0154】なお、上述の例では、入力部11の構成について入力デバイス処理部111、入力モード処理部112、入力メディア処理部113の3つのモジュールを持ち、出力部14についても出力メディア処理部141、出力モード処理部142、出力デバイス処理部143の3つのモジュールを持つことを前提として記述しているが、本装置を実現する入出力処理部品の構成はこれに限定されるものではなく、種々の変形を行うことが可能である。

【0155】また、上述の例では、入力部と出力部が独立した別の構成要素として記述しているが、これらを一体化した入出力部として実現することも可能であり、そのような実現方法も本発明の趣旨の範囲内である。

【0156】また、上述した例では、メディア割当決定部を対話管理部や問題解決手段と独立した構成部品として、また、問題解決手段についても同様に入力部や出力部、対話管理部等に対して独立した部品として記述しているが、本装置を実現する部品の構成方法はこれに限定されるものではなく、例えば、メディア割当決定機能を持つ対話管理部を実現することで独立した部品としてのメディア割当決定部がなくとも本装置を実現することが可能であり、また、例えば、問題解決手段が対話管理部の機能やメディア割当決定部の機能、さらには入力部の

機能や出力部の機能を持つように構成することも可能であり、そのような実現方法も本発明の趣旨の範囲内である。

【0157】このように、本発明を実現する装置はその部品構成について種々の実現方法があり、ここに説明した部品の構成に限定されるものではないことを断つておく。また、上述した例では、メディア割当決定部12の制御方式について、ある時点におけるメディア割当と、その時点での状況に基づいて次のメディア割当を決定する次メディア割当選択テーブル122を用いることを前提として記述しているが、本装置を実現するメディア割当決定方式はこれに限定されるものではなく、例えば状態遷移モデルに基づくメディア割当決定方式においても実現可能であり、また、例えば次メディア割当選択テーブル122に選好度などのスロットや情報を付け加えることで、よりきめ細かいメディア割当の選択を実現することも可能になる。

【0158】また上述した例ではメディア割当生成部121においてメディア割当を決定する際に用いるメディア割当のヒストリ情報として1回分、すなわち現在のメディア割当情報のみ明確に記述されているが、例えば次メディア割当選択テーブル122の条件部に過去のメディア割当の状態を記述することにより、メディア割当の決定に対し、メディア割当のヒストリ情報を利用することができる。

【0159】また、上述の例では、階層プラン生成方式に基づく対話管理を前提として記述するが、本装置を実現する対話管理方式はこれに限定されるものではなく、例えば状態遷移モデルに基づく対話制御方式においても実現可能である。

【0160】更に、上述の例では、利用者が利用可能な入力方法をディスプレイ上に他の情報と共に表示することにより利用者に知らしめているが、本装置を実現する入力メディアの提示形式はこれに限定されるものではなく、例えばもう一つのディスプレイに表示するなど他の出力デバイスから提示することとしても実現可能である。

【0161】また、上述の例では、利用者が利用可能な入力方法のみをディスプレイに表示することで利用者に知らしめているが、利用者に知らしめるのは入力方法に限定されるものではなく、出力方法についても同様に表示することが可能であり、例えば耳の不自由な利用者に対して音声による応答を本装置が行ったとしても、利用者がそれを知ることができるため、利用者がメディア割当の変更を要求することが可能となる。

【0162】また、更に例えば、メディア割当決定部に特定利用者モデル記憶部を設け、そこに利用者が利用可能なメディアもしくは利用不可能なメディアを予め記録しておくことで、上記の耳の不自由な利用者に対する場合においても利用者が毎回要求することなしに、個々の

利用者に対応して最適なメディア割当を設定することが可能になる。

【0163】また、例えば、上記のようなメディア割当に関する利用者個々の情報を、例えば、ICカードのような可変媒体に記録しておき、例えば、ICカード読取り可能なデバイスを本装置に装備して、このデバイスを利用して利用者が利用開始時にそのICカードを挿入して読み取らせ、特定利用者モデル記憶部にその読み取り情報を記憶させることで、不特定多数を対象とした状態においても、予め利用者個々の情報を設定して対処することができるようになり、その結果、利用者個々の事情に応じたメディア割当の設定が利用開始の最初の時点から可能になる装置が得られるようになる。この場合、個々の利用者の最適なメディア割当が利用開始の最初の時点から実施されるので、変更の要求操作をすることなしに、円滑な利用が可能になることからサービス向上に役立つ。

【0164】また、特定利用者のみが利用するようなシステムの場合には例えば、不揮発性メモリに特定利用者モデル記憶を行うようにし、利用者が更新できるようにしたり、学習機能を持たせて記憶内容を学習結果を反映させるようにする等のことによって、利用を重ねるに連れ、その利用者が利用しやすいインターフェイスを提供することができるようになる。

【0165】さらにメディア割当決定部にID3やニューロ等の機能的な学習機能を装備し、特定利用者モデル記憶部の内容を学習により自動更新することにより、利用者が自分で利用者モデルを更新せずとも、利用を重ねるだけで次第にその利用者が利用し易いインターフェイスを提供することができるようになる。

【0166】また、以上では特定利用者モデル記憶部の記憶内容について、特定の利用者が利用可能なメディアもしくは利用不可能なメディアとしたが、特定利用者モデル記憶部の記憶内容はこれに限定されるものではなく、例えば、特定利用者が選択するメディアについても同様に、特定利用者モデル記憶部に記憶することで、より個々の利用者にとり好みのインターフェイスを提供することができるようになり、そのような実現方法も本発明の趣旨の範囲内である。

【0167】さらに利用者が利用可能な入り方を知らしめるための出力メディア自体をメディア割当決定部において決定し、その決定に基づく出力メディア割当を用いて出力することで、利用者に知らしめるといった方法を採用することも可能であり、そのような実現方法も本発明の趣旨の範囲内である。

【0168】なお、上述の例では、利用者の読み間違いによる音声認識の失敗を扱う例を示したが、例えば、環境音（ノイズ）が原因となって音声入力が阻害される場合等のように、利用するメディアの状況の変化によって対話に影響が発生するような状況においても、入力及び

出力メディアの状態を監視するメディア監視機能をメディア切替部に組み込み、ある時点でメディア切替部によって設定され、現在利用しようとしているメディア割当が、障害なく利用され、利用者と対話装置とのコミュニケーションが無事に達成されたかを監視することによって、障害解消のための処理を行うことが可能となる。

【0169】すなわち、例えば、利用者からの音声入力を受けようとしている場面や、利用者への出力を音声によって行おうとしている場面等において、例えば、周囲騒音のレベル等を常時監視することによって、例えば、大きな騒音等の発生による利用者と対話装置とのコミュニケーションの障害の生じた場合に、これを対話管理部に通知して例えば、音声による入出力をやめて、画面表示とキーボード入力を行わせるようにする等、何等かの障害発生の場合に、適切にメディア変更を行ったり、あるいは、問い合わせ等の障害解消のためのサブ対話を起動させる等の処理を実現することが可能になり、その結果、環境等の変化に対しても対処し得る対話装置が実現できる。このように本発明においては、その趣旨を選択しない範囲で様々な変形を行うことが可能である。

【0170】

【発明の効果】以上説明したように、本発明によれば、対話の状況及びその時点で利用可能な入出力手段に合せて適切な入出力メディアを動的に選択するようにしたことで、状況に応じて使いやすいインターフェイスを利用者に提供することが可能となる。また、利用者がどのような手段で入力を行うことが出来るかを容易に知ることができ、より使い易いインターフェイスを提供することが可能となる。

【0171】また、アプリケーションが出力を要求する情報の質と置に対応して利用者が受け入れ易くわかるに易い出力手段の組み合わせを設定することが可能となり、利用者が要求する入出力方法を用いることでより個々の利用者に使いやすいインターフェイスを提供することが可能となり、利用者にとって好ましい入出力手段を用いることで、より個々の利用者に使い易いインターフェイスを提供することが可能となる。更には、入出力手段を変更することで利用者との対話の失敗から回復し、利用者との対話を継続することが可能となるなどの実用上多大な効果が奏せられる。

【図面の簡単な説明】

【図1】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのプロック構成図。

【図2】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するための対話規則の例を示す図。

【図3】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するための対話プリミティブ

記憶テーブル例を示す図。

【図4】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのプラン記憶部(起動時)の例を示す図。

【図5】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのプラン記憶部(待機時)の例を示す図。

【図6】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するための次メディア割当選択テーブルの例を示す図。

【図7】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのモードメディア割当対応表の例を示す図。

【図8】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するための想定する動作例シナリオの例を図。

【図9】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのプラン記憶部(初期条件入力要求時)の例を示す図。

【図10】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのプラン記憶部(初期換算終了時、抜粋)の例を示す図。

【図11】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのプラン記憶部(条件絞り込み対話、抜粋)の例を示す図。

【図12】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのプラン記憶部(利用者割込時、抜粋)の例を示す図。

【図13】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのプラン記憶部(利用者割込対処時、抜粋)の例を示す図。

【図14】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのプラン記憶部(音声認識失敗への対処、抜粋)の例を示す図。

【図15】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するための詳細な構成図。

【図16】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのメディア割当“ヒューマン”的画面例を示す図。

【図17】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのメディア割当“変換表示音入出力”的画面例を示す図。

【図18】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのメディア割当“タッチパネル表示選択、要素名音声読み上げ”的画面例を示す図。

【図19】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するための対話管理のアルゴリズム例を示す図。

50 【図20】本発明を説明するための図であって、本発明

の具体的な実施の態様を説明するための待機状態例を示す図。

【図21】本発明を説明するための図であって、本発明の具体的な実施の態様を説明するためのメディア割当「表形式タッチパネル表示選択・検索結果件数音声出力」の画面例を示す図。

【符号の説明】

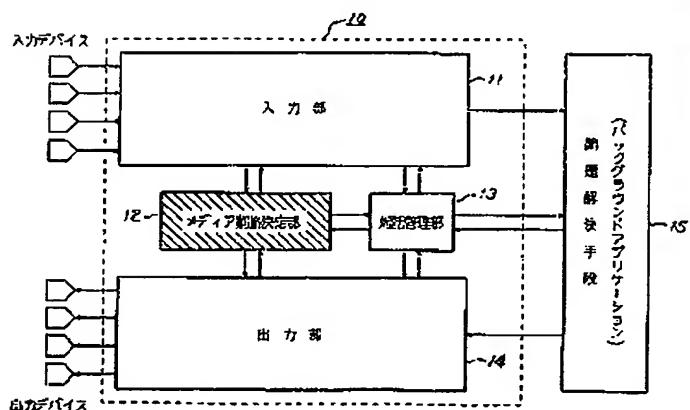
- 10…マルチモーダル対話装置
- 11…入力部
- 12…メディア割当決定部
- 13…対話管理部
- 14…出力部
- 111…入力デバイス処理部
- 112…入力モード処理部
- 113…入力メディア処理部
- 121…メディアアロケーション生成部

- \* 122…次メディア割当選択テーブル
- 123…現メディア割当記憶スロット
- 124…モードメディア割当対応表
- 131…対話プランナ
- 132…対話規則テーブル
- 133…対話プリミティブ記憶テーブル
- 141…出力メディア処理部
- 142…出力モード処理部
- 143…出力デバイス処理部
- 143a…出力メディア振り分けスイッチ
- 143b…出力メディア処理部
- 143c…メディア・モードインターフェイス
- SW…メディア選択スイッチ
- 1DVa～1DVn…入力デバイス
- ODVa～ODVn…出力デバイス

\*

【図1】

【図16】



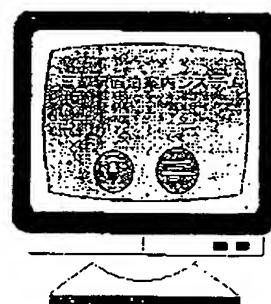
メディア選択「ヒューマン」の画面例

【図3】

【図20】

ファクトID	目標	仕様
f01	MA問合せ(条件)	メディアアロケータ呼び出し(条件)
f02	検出特徴	
f03	MA設定(MA)	
f04	変数入力(変数)	入力メディア処理部呼び出し(入力要求(変数))

対話プリミティブ記憶テーブル



特許文献

【図2】

ルールID	目標	摘要条件	サブ目標リスト
r00	初期化		ユーザ横笛、MA選択(初期)、開始化マシン、 タスク実行、終了セッション
r01	タスク実行		初期条件入力、探索、候補決定、候補選択
r02	初期条件入力		入力要求 (A→D、入力変数リスト) MA選択 (入力要求 (入力変数リスト)) 入力処理 (入力変数リスト)。
r03	開始セッション		開始要求 (U→S)、開始確認 (S→U)
r04	結合セッション		開始要求 (S→U)、開始確認 (U→S)
r05	終了セッション		終了要求 (U→S)、終了確認 (S→U)
r06	終了セッション		終了確認 (S→U)
r07	候補決定	候補数=1	選択要求
r08	候補決定		候補条件取り込み、探索、候補決定
r09	候補決定		MA選択、選択肢表示、ユーザ選択
r10	選択要求		MA選択 (特機)、横笛特機
r11	ユーザ横笛		MA選択 (特機)、横笛特機
r12	MA選択 (条件)		MA選択 (条件)、MA選択 (MA1)、 MA選択 (データ)
r13	表現形式変更要求 (U→S)		MA選択 (条件)、MA選択 (MA1)、 MA選択、データ出力
r14	探索条件取り込み		未入力成員条件 (D→A、未入力変数リスト)、 number (入力変数、入力変数リスト)、 追加候補入力 (入力変数、候補数込み)、 MA選択 (ユーザ要求 (MA))、選択肢表示、 ユーザ選択
r15	選択要求	WA要求 (MA, U→S)	MA選択 (ユーザ要求 (MA))、選択肢表示、 ユーザ選択
r16	選択要求	成員条件変更要求 (条件)	MA選択 ()、 MA選択 ( )
r17	追加候補入力 (入力変数、入力目的)		MA選択 (追加候補入力 (入力変数、入力目的))、 入力目的表示 (D→U、追加候補入力 (入力目的))、 入力処理 (入力変数)
r18	入力処理 (入力変数)		変数入力 (入力変数)
r19	入力処理 (入力変数)		再入力要求、 変数入力 (入力変数)
r20	入力処理 (入力変数) 再入力要求		MA選択 (入力実現 (入力変数))、 変数入力 (入力変数)

対話規則

【図4】

ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01,p02,p03,p04,p05]	
p01	ユーザ検出	処理中	[ ]	
p02	MA選択(初期)	未処理		
p03	開始セッション	未処理		
p04	タスク実行	未処理		
p05	終了セッション	未処理		

プラン記憶部(起動時)

【図5】

ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01,p02,p03,p04,p05]	
p01	ユーザ検出	処理中	[p11,p12]	
p02	MA選択(初期)	未処理		
p03	開始セッション	未処理		
p04	タスク実行	未処理		
p05	終了セッション	未処理		
p11	MA選択(待機)	成功	[p111,p112]	待機
p12	検出待機	処理中	[ ]	
p111	MA選択(待機,MA1)	成功	[ ]	MA1=待機
p112	MA設定(待機)	成功	[ ]	

プラン記憶部(待機時)

[図6]

ノードID	現メディア割当	条件	次メディア割当
m01	---	呼出条件=待機	待機
m02	---	呼出条件=初期	ヒューマン
m03	---	呼出条件=入力要求	次音声入力画面音声出力
m04	---	呼出条件=選択要求 & 要素数<=4	タッチパネル表示選択・要素名表示読み上げ
m05	---	呼出条件=選択要求 & 5<=要素数<=100	表形式タッチパネル表示選択・候補表示要素数音声出力
m06	( 音声基底, XXX )	呼出条件=入力失敗	( ペン文字列, XXX )

次メディア割当選択テーブル

[図7]

ノードID	メディア割当名	制約	入力モード組
m001	ヒューマン	音声入出力可能	( 音声音源, 音声音源+顔画像 )
m002	待機	入力センサ利用可能	( センサ検出待機, 待機メッセージ )
m003	次音声入力画面音声出力	実数型E(DATE, INT, MONEY) & 入力次数数<=4 & 音声入力可能	( 音声音語, 音声音語+次数画面表示 )
m004	ペン文字列入力・ヒューマン 対応	音声出力可能	( ペン文字列, 音声音源+顔画像 )
m005	タッチパネル選択・要素名 表示読み上げ	要素数=N	( N要素タッチパネル(並列), 音声音語 ( N件 ) + N要素並列配置 )
m006	表形式タッチパネル表示選択 ・候補表示要素数音声出力	要素数=N	( N要素タッチパネル(表形式), 音声音語 ( N件 ) + N要素表示出力 )

モード・ディア割当対応表

【図8】

A)	案内システムの前に一人の銀行者が立つ センサが反応してディスプレイがオンになり、システムのタイトル画面 効果音とともに入力を促すプロンプトが出る。
B)	銀行者の音声を使ってシステムに働きかける 利用者 (U2): 「杏を探してください」 システム (S1): 「了解しました」
C)	初期データ登録会話 (S2): 「いつからご宿泊ですか？」 (U2): 「今日からお願ひします」 (S3): 「ご出迎えいつになりますか？」 (U3): 「あさっての予定です」 (S4): 「何部屋が必要ですか？」 (U4): 「一人ででの一部屋で結構です」 (S5): 「一泊あたりのご予算はいくらくらいですか？」 「画面に表示選択肢の中から選んで下さい」 (U5): 「6000円以下でお願いします」 (U6): 「一泊6000円程度までお宿に本日から12日水曜日までの3泊、 一泊室のご予約を御希望ですね。」 (U7): 「では、ただ今から検索しますのでしばらくお待ち下さい。」
D)	鍵匙の表示 (S8): 「お待たせいたしました」 (S9): 「おっしゃった条件で本日から宿泊可能な宿泊施設の 施設A室、ひなびた庵、奥山院、の3件です。」 (S10): 「各宿泊施設の詳しい情報を御観になりたい場合は、その名前をしゃべるか、 画面の該当部分を指で押して下さい。」
E)	利用者主導の会話 (U6): 「それぞれ道順と時間を教えて下さい」 (S11): 「奥山院までは約1時間半ほどです。 ひなびた庵には列車で30分ほど行ったところから バスに乗り換えて30分ほどかかります。 奥山院はここからバスで30分ほど乗ってから 30分ほど歩いたところにあります」
F)	検索絞り込みの提案 (S13): 「件数がS34とまだ多いので、条件を絞り込みたいと思います」 (S14): 「どのようなところを観光される予定でしょうか？」
G)	ユーザーの読み間違い(音声認識失敗)への対処 (U8): 「りゅうあんじです」 (S15): 「は？もう一度お頼りします」 (U9): 「りゅうあんじです」 (S16): 「認識できませんでした。ベンで行き先の名前を書いてください」 (S17): 「龍安寺(りょうあんじ)ですね。検索しますのでしばらくお待ち下さい」

想定する動作別シナリオ

【図9】

ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01, p02, p03, p04, p05]	
p01	ユーザ検出	成功	[p11, p12]	
p02	MA選択(初期)	成功	[p11, p12]	
p03	開始セッション	成功	[p31, p32]	
p04	タスク実行	成功	[p41, p42, p43, p44]	失敗
p05	終了セッション	失敗		
p11	MA選択(初期)	成功	[p111, p112]	MA1=初期
p12	共用情報	成功	[]	
p111	MA組合せ(初期, MA1)	成功	[]	
p112	MA組合せ(初期)	成功	[]	MA2=ヒューマン
p21	MA組合せ(初期, MA2)	成功	[]	
p22	MA組合せ(ヒューマン)	成功	[]	
p31	開始表示( U→S)	成功	[]	
p32	開始確認( S→U)	成功	[]	
p41	初回条件入力	処理中	[p411, p412, p413]	
p42	状態	未処理		
p43	候選決定	未処理		
p44	候選確認	未処理		
p411	人選択(A-D)【選択日, 選択日, 人選, 不採】	成功	[]	
p412	始動表示(初期)【選択日, 選択日, 人選, 不採】, MA3	成功	[p4111, p4112]	
p412	人選選択【選択日, 選択日, 人選, 不採】	処理中		
p4111	MA組合せ(人選表示)【選択日, 選択日, 人選, 不採】	成功		
p4112	MA組合せ(人選表示)【選択日, 選択日, 人選, 不採】	成功		MA1=候選表示入出力

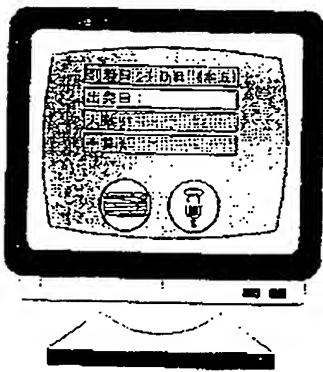
プラン記憶部(初期条件入力要求時)

【図10】

ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01,p02,p03,p04,p05]	
p01	ユーザ検索	成功	[p11,p12]	
p02	MA選択(初期)	成功	[p21,p22]	
p03	開始セッション	成功	[p31,p32]	
p04	タスク実行	処理中	[p41,p42,p43,p44]	
p05	終了セッション	未処理		
…省略…				
p41	初期条件入力	成功	[p411,p412,p413]	
p42	検索	成功	[ ]	
p43	候補決定	処理中	[p431]	
p44	候補確認	未処理		
…省略…				
p431	選択要求	処理中	[p4311,p4312,p4313]	
p4311	MA選択(選択表示)	処理中	[p43111,p43112]	
p4312	タスク実行	未処理		
p4313	終了セッション	未処理		
p43111	MA選択(初期)	処理中	[ ]	
p43112	開始セッション	未処理		

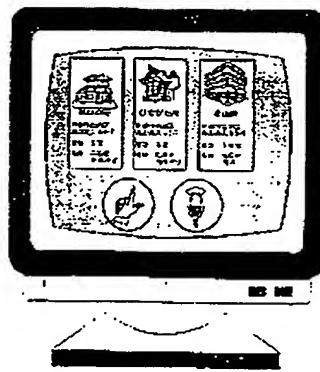
プラン記憶部(初期検索終了時,抜粋)

【図17】



メモリ表示・登録登出登録入出力画面例

【図18】



ノック表示・タッチパネル表示登録登出登録入出力画面例

[図11]

ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01,p02,p03,p04,p05]	
p01	ユーザ検出	成功	[p11,p12]	
p02	MA選択(初期)	成功	[p21,p22]	
p03	開始セッション	成功	[p31,p32]	
p04	タスク実行	処理中	[p41,p42,p43,p44]	
p05	終了セッション	未処理		
…略…				
p41	初期条件入力	成功	[p411,p412,p413]	
p42	検索	成功	[ ]	
p43	候補決定	失敗	[p431]	
p44	候補選択	未処理		
…略…				
p431	選択要求	失敗	[p4311,p4312,p4313]	
p4311	MA選択(選択肢表示)	失敗	[p43111,p43112]	
p4312	選択肢表示	未処理		
p4313	ユーザ選択	未処理		
p43111	MA開合(選択肢表示,MA1)	失敗	[ ]	
p43112	MA設定(MA1)	未処理		
p43	候補決定	再実行	[p4301,p4302,p4303]	
p4301	検索条件取り込み	処理中	[p43011,p43012,p43013]	
p4302	検索	未処理		
p4303	候補決定	未処理		
p43011	未入力選択開合( D→A, 入力候補リスト )	成功	[ ]	候補リスト=[候補1,候補2,候補3]
p43012	member(入力候補, [ 要素予定地, 検査式, 候補表示式 ] )	成功		候補=候補3
p43013	選択条件入力( 要素予定地, [ 検索取り込み ] )	処理中	[p4301311,p4301312]	
p430131	MA選択(選択条件入力( 要素予定地, [ 検索取り込み ] ))	成功	[p4301311,p4301312]	
p430132	入力目的表示( C~U, [ 要素予定地, 検索取り込み ] )	成功		
p430133	入力検査( 要素予定地 )	処理中	[p4301331]	
p4301331	検索入力( 要素予定地 )	処理中	[ ]	
…略…				

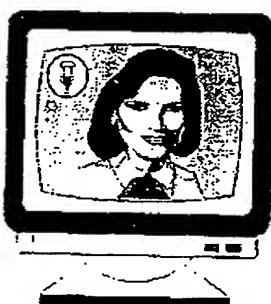
プラン記憶部(条件取り込み実施,抜粋)

[図12]

ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01,p02,p03,p04,p05]	
p01	ユーザ検出	成功	[p11,p12]	
p02	MA選択(初期)	成功	[p21,p22]	
p03	開始セッション	成功	[p31,p32]	
p04	タスク実行	処理中	[p41,p42,p43,p44]	
p05	終了セッション	未処理		
…中略…				
p41	初期条件入力	成功	[p411,p412,p413]	
p42	検索	成功	[]	
p43	候補決定	処理中	[p431]	
p44	候補確認	未処理		
…中略…				
p431	選択要求	処理中	[p4311,p4312,p4313]	
p4311	MA選択(選択提示)	成功	[p43111,p43112]	
p4312	選択肢提示	成功	[]	
p4313	ユーザ選択	未処理	[]	
…中略…				

プラン記憶部(利用者割込時抜粋)

[図21]



メモリモード表示式ターミナル画面表示 検索結果件数表示用表示例

[図13]

コードID	目標	状態	状況	アクションリスト	結果
p0	対話セッション			[p01, p02, p03, p04, p05]	
p01	ユーザ提出 MA登録(初期)			成功 [p11, p12]	
p02	監査セッション			成功 [p21, p22]	
p03	タスク実行			成功 [p31, p32]	
p04	統合セッション			処理中 [p41, p42, p43, p44]	
p05					
...中略...					
p11	初期条件入力			成功 [ ]	
p12	検索 候補表示			処理中 [p431]	
p43	候補表示				
p431	MA登録(初期表示)			成功 [p4311, p4312, p4313]	
p4311	MA登録(初期表示)			失敗 [ ]	
p4312	最初表示			失敗 [ ]	
p4313	ユーザ登録			失敗 [ ]	
...中略...					
p431	最初表示				
p43101	MA登録(ユーザ登録(出力)([選択, 初期登録]))			成功 [p43101, p43102, p43103]	
p43102	最初表示			成功 [ ]	
p43103	2-次登録			処理	
p431011	MA登録(ユーザ登録(出力)([選択, 初期登録]))			成功 [ ]	
p431012	MA登録(MA1)			成功 [ ]	
p431012	MA登録(MA1)				

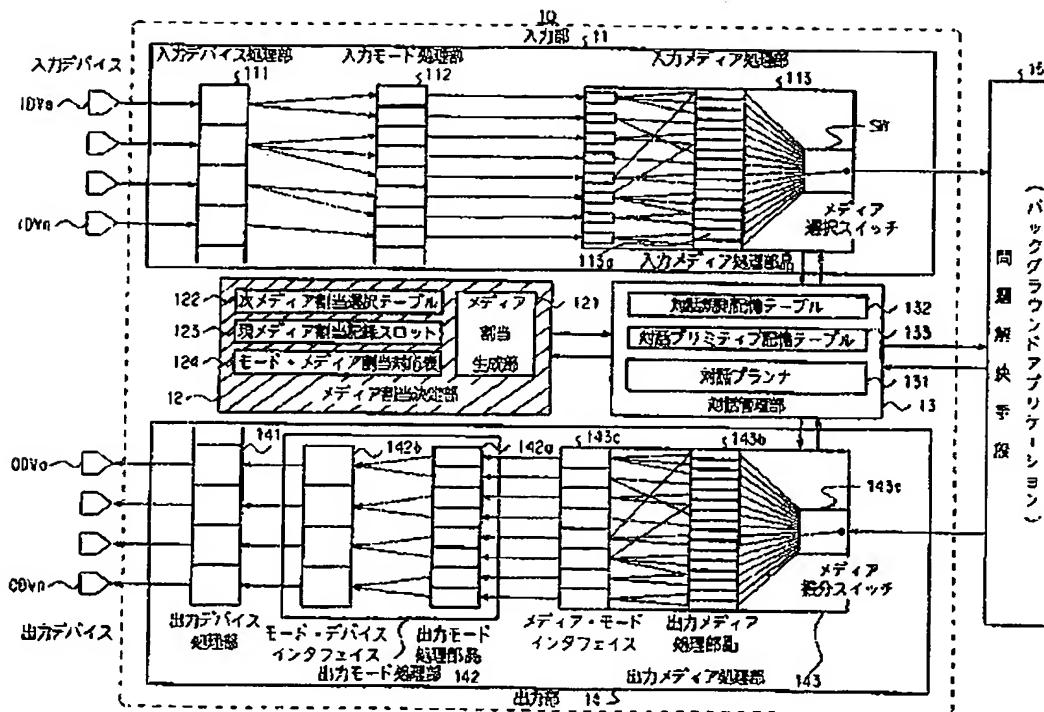
プラン記憶部(利用者割込判定時、操作)

[図14]

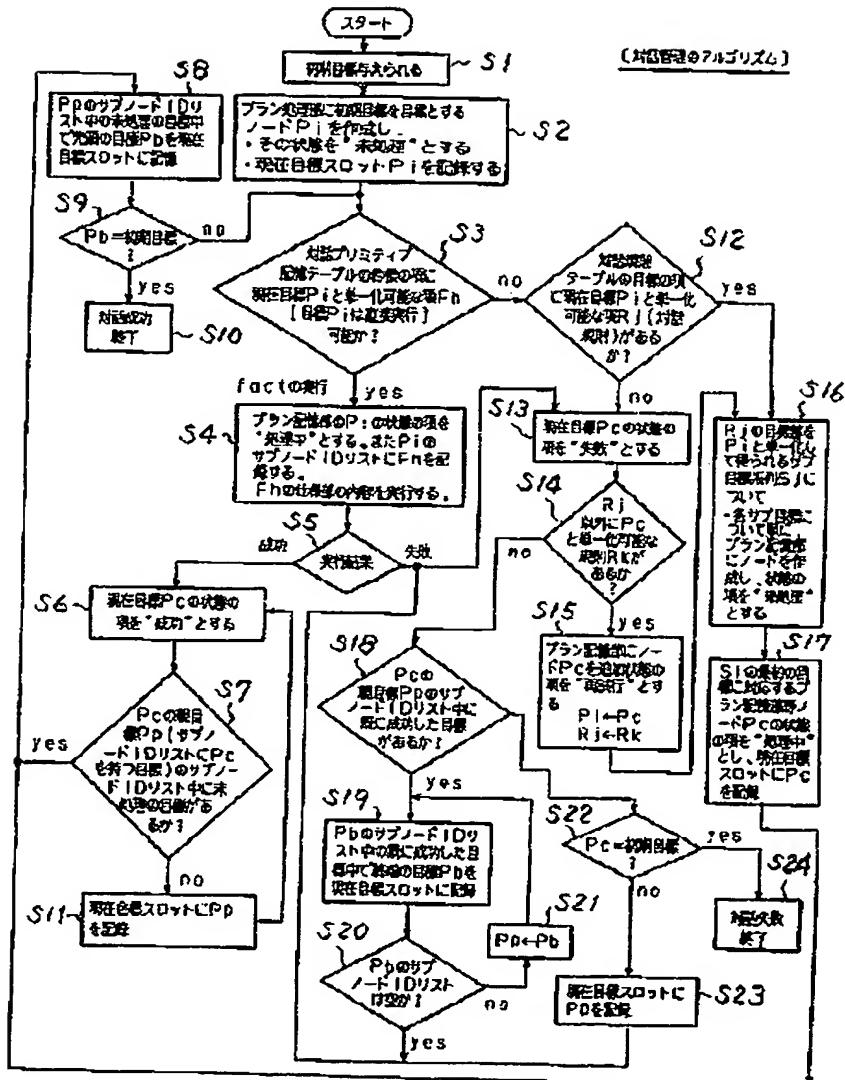
ノードID	名前	状態	サブノードIDリスト	結果
P0	対話セッション	処理中	[P01, P02, P03, P04, P05]	
P01	ユーチューブ検索	成功	[P11, P12]	
P02	MA選択(初期)	成功	[P21, P22]	
P03	開始セッション	成功	[P31, P32]	
P04	タスク実行	処理中	[P41, P42, P43, P44]	
P05	終了セッション	未処理		
-中略-				
P41	初期操作入力	成功	[P411, P412, P413]	
P42	検索	成功	[1]	
P43	候補決定	失敗	[P431]	
P44	候補選択	未処理		
-中略-				
P431	選択要求	失敗	[P4311, P4312, P4313]	
P4311	MA選択(選択肢表示)	失敗	[P43111, P43112]	
P4312	選択肢表示	未処理		
P4313	ユーチューブ検索	未処理		
P43111	MA結合(選択肢表示, MA1)	失敗	[1]	
P43112	MA設定(MA1)	未処理		
P43	候補決定	再実行	[P4301, P4302, P4303]	
P4301	候補条件取り込み	処理中	[P43011, P43012, P43013]	
P4302	検索	未処理		
P4303	候補決定	未処理		
P43011	未入力変数結合(D→A, 未入力変数リスト)	成功	[1]	未入力変数→{検索用, 候補, 選択肢} A200-候補数
P43012	member(入力変数, [観光予定地, 選択肢, 候補表示式])	成功	[1]	
P43013	未入力変数入力(観光予定地, [候補表示式])	処理中	[P4301311, P4301312]	
P430131	MA選択(追加条件入力(観光予定地, [候補表示式]))	成功	[P4301311, P4301312]	
P430132	入力目的表示(D→I, [追加条件入力, 候補表示式])	成功	[1]	
P430133	入力処理(観光予定地)	失敗	[P4301331]	
P4301331	変数入力(観光予定地)	失敗	[1]	
-中略-				
P430133	入力処理	失敗	[P43013301, P43013302]	
P43013301	最大値要求	失敗	[1]	
P43013302	変数入力(観光予定地)	失敗	[1]	
P430133	入力処理(観光予定地)	再実行	[P43013311, P43013312, P43013313]	
P43013311	MA選択(入力失敗(観光予定地))	処理中	[P430133111, P430133112]	
P43013312	再入力要求	未処理		
P43013313	変数入力(観光予定地)	未処理		
P430133111	MA結合(入力失敗(観光予定地), MA1)	処理中	[1]	
P430133112	MA設定(MA1)	未処理		

プラン記憶部(音声記憶失敗への対応, 終了)

[図15]



[図19]



フロントページの続き

(72)発明者 田中 克己

大阪府大阪市北区大淀中1丁目1番30号  
株式会社京芝関西支社内

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-091112  
(43)Date of publication of application : 04.04.1997

---

(51)Int.Cl. G06F 3/14  
G06F 3/02  
G10L 3/00  
G10L 3/00

---

(21)Application number : 08- 056613 (71)Applicant : TOSHIBA CORP

(22)Date of filing : 14.03.1996 (72)Inventor : KONO YASUYUKI  
IKEDA TOMOO  
CHINO TETSURO  
TANAKA KATSUMI

---

(30)Priority

Priority 07205202 Priority 19.07.1995 Priority JP  
number : date : country :

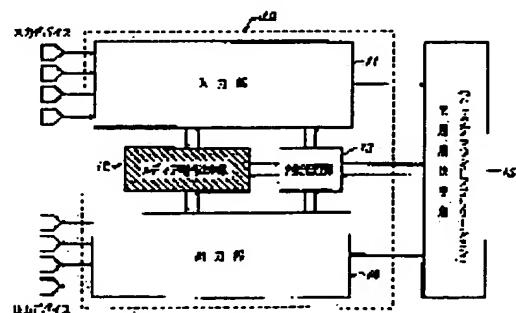
---

## (54) MULTIMODAL INTERACTIVE DEVICE AND INTERACTIVE METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To enable interaction by dynamically switching input/output media in response to situations.

SOLUTION: This device is provided with an input means 11 for interpreting the input data of a user from one device corresponding to input medium allocation among different devices for various kinds of inputs and applying that input corresponding to interpretation to a problem solving means 15, problem solving means 15 for outputting the result by performing processing corresponding to the purpose



of a system corresponding to this input, output means 14 for interpreting the output of this problem solving means, distributing this interpreted result to any device for output corresponding to output medium allocation among different devices for various kinds of outputs and outputting the result to the user, and medium allocating means 12 for dynamically generating the method for selecting input data and interpreting them and the method for combining those data corresponding to the condition of interaction with the user by the input means 11 and the output means 14 in order to construct the input data to the problem solving means from the input data applied from the user to the plural input devices.

---

## LEGAL STATUS

[Date of request for examination] 14.09.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3729918

[Date of registration] 14.10.2005

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

\* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

## CLAIMS

---

### [Claim(s)]

[Claim 1] In the dialogue equipment which performs necessary processing by the dialogue with a user, and outputs a result An input means to interpret the input from at least one device for an input according to input media allocation among the various devices for an input with which it differs for acquiring the information from a user, and to generate response information based on the interpretation, A problem-solving means to carry out processing according to the necessary object to each of this generated response information, and to output a result to it, An output means to output the result of having been based on the output of said problem-solving means, from the device for an output which followed output media allocation among various different devices for an output, A means to choose said response information given from said input means to said problem-solving means from said response information generated for said every input media, Said input media allocation which assigns at least one of the means which combine the response information generated for said every input media, And said output media allocation which assigns at least one of said problem-solving means to the means which combine the means and them which choose the information from this problem-solving means in order to acquire the print-out outputted to an output means, Multi-modal dialogue equipment characterized by providing a media allocation means for it to be alike, and for it to be based and to perform input or output media allocation according to a dialogue situation with said user.

[Claim 2] In the dialogue equipment which performs necessary processing by the dialogue with a user, and outputs a result An input means to interpret the input from at least one device for an input according to I/O media allocation among the various devices for an input with which it differs for acquiring the information from a user, and to generate response information based on the interpretation, A problem-solving means to carry out processing according to the necessary object to each of this generated response information, and to output a result to it, An output means to output the result of having been based on the output of said problem-solving means, from the device for an output which followed I/O media allocation among various different devices for an output, A means to choose said response information given from said

input means to said problem-solving means from said response information generated for said every input media, At least one of the means which combine the response information generated for said every input media And at least one of said problem-solving means to the means which combine the means and them which choose the information from this problem-solving means in order to acquire the print-out outputted to an output means Multi-modal dialogue equipment characterized by providing the media allocation means which assigns I/O media according to a dialogue situation with said user based on the I/O media allocation which \*\*\*\*\* is made to correspond and is assigned.

[Claim 3] In the dialogue equipment which performs necessary processing by the dialogue with a user, and outputs a result An input means to interpret the input from at least one device for an input according to I/O media allocation among the various devices for an input with which it differs for acquiring the information from a user, and to generate response information based on the interpretation, A problem-solving means to carry out processing according to the necessary object to each of this generated response information, and to output a result to it, An output means to output the result of having been based on the output of said problem-solving means, from the device for an output which followed I/O media allocation among various different devices for an output, The dialogue management tool which directs modification processing of I/O media allocation according to the processing result from said problem-solving means, and manages the I/O by said input means and output means, In order to obtain the input from the input data given from the user to at least one input device to said problem-solving means In order to obtain the expression directed to a user from the output of the at least one and said problem-solving means of the combination of the input approach of the device and input data which are used among said devices for an input according to a predetermined regulation While carrying out I/O media allocation for making the both sides of at least 1\*\* of the combination of the output method of a device and output data which use among said devices for an output according to a predetermined regulation correspond, choosing, and assigning It is multi-modal dialogue equipment characterized by providing the media allocation means which processes the modification in order to direct modification processing of I/O media allocation, and carries out new media allocation at the time of a carrier beam.

[Claim 4] A problem-solving means to carry out computation according to the object of a system to an input response, and to output a result to

it, An input means to interpret a user's input data from at least one device among various different devices for an input corresponding to I/O media allocation, and to give the input corresponding to an interpretation to said problem-solving means, An output means to obtain the processing result from said problem-solving means, to distribute this to the device for an output corresponding to I/O media allocation among various different devices for an output, and to output to a user, The dialogue management tool which directs modification processing of I/O media allocation according to the processing result from said problem-solving means, and manages the I/O by said input means and output means, A means to choose the input given to said problem-solving means from said response information generated for every input media, And at least one of the means which combine the response information generated for every input media And in order to generate the expression directed to a user from the output of said problem-solving means, while choosing said I/O media allocation which is the combination of the both sides of at least 1\*\* of the means which combines the means and them which choose the output It is multi-modal dialogue equipment characterized by having a media allocation means to process the modification in order to direct modification processing of I/O media allocation, and to choose new I/O media allocation, at the time of a carrier beam.

[Claim 5] In said dialogue management tool, one [ at least ] situation of an input unit and an output means of operation is supervised suitably. When the failure over utilization of the I/O media allocation which current selection is made and is used occurs, while having the media monitoring function which notifies the fault information showing generating of this failure to a media assignment-decision means Claims 2 or 3 characterized by having the function to change I/O media allocation suitably if this fault information is received in said media allocation means, or multi-modal dialogue equipment given in 4 any 1 terms.

[Claim 6] When information required for said problem-solving means for the problem solving of a system needs to be acquired from a user the input request function for which it asks so that input request information may be generated to said dialogue management tool and I may have a user input said input required information -- or Information required for said dialogue management tool since the output to a user is constituted according to the I/O media allocation to which said output means is set the output request function for which it asks so that the output request

information which asks for an output may be generated for a problem-solving means and a problem-solving means may output said output required information to it, and \*\*, while having at least one function inside For said media allocation means, directions of I/O media allocation modification processing at the time of a carrier beam Claim 3 characterized by having the function to perform new media allocation using at least one of said input request information or said output request information, or multi-modal dialogue equipment given in 4 any 1 terms.

[Claim 7] Interpret the information acquired from the user by the dialogue with a user, and the information corresponding to an input is generated. The information corresponding to an input is sent to the problem-solving equipment which carries out processing according to the necessary object to each of this information corresponding to the generated input, and outputs a result to it through means of communications. In the dialogue equipment which outputs the processing result of this information corresponding to an input acquired from problem-solving equipment through means of communications by the dialogue with a user An input means to interpret the input from at least one device for an input according to I/O media allocation among the various devices for an input with which it differs for acquiring the information from a user, and to generate response information based on the interpretation, An output means to output the result obtained from said problem-solving equipment through means of communications from the device for an output which followed I/O media allocation among various different devices for an output, Said sending corresponding to an input generated in the input means to said problem-solving equipment, and the means of communications which performs the receipt of the problem-solving result which problem-solving equipment generated and was outputted, A means to choose said response information sent to said problem-solving equipment from said input means from said response information generated for said every input media, At least one of the means which combine the response information generated for said every input media And at least one of the means which combine the means and them which choose the information from this problem-solving equipment in order to acquire the print-out outputted to an output means from said problem-solving equipment Multi-modal dialogue equipment characterized by providing a means for the media allocation which performs I/O media allocation for generating the I/O media allocation which is the combination of \*\*\*\*\* according to a dialogue

situation with said user, and performing a dialogue with a user.

[Claim 8] Interpret the information acquired from the user by the dialogue with a user, and the information corresponding to an input is generated. The information corresponding to an input is sent to the problem-solving equipment which carries out processing according to the necessary object to each of this information corresponding to the generated input, and outputs a result to it. In the dialogue equipment which outputs the processing result of this information corresponding to an input acquired from problem-solving equipment by the dialogue with a user An input means to interpret the input from at least one device for an input according to I/O media allocation among the various devices for an input with which it differs for acquiring the information from a user, and to generate response information based on the interpretation, An output means to output the result obtained from said problem-solving equipment through means of communications from the device for an output which followed I/O media allocation among various different devices for an output, The dialogue management tool which directs modification processing of input media allocation and output media allocation according to the processing result from said problem-solving equipment, and manages the I/O by said input means and output means, Said sending corresponding to an input generated in the input means to said problem-solving equipment, and the means of communications which performs reception of the problem-solving result which problem-solving equipment generated and was outputted, In order to acquire the information corresponding to the input given to said problem-solving equipment from the input data given from the user to at least one input device In order to obtain the expression shown to a user from the result obtained from at least one and said problem-solving equipment of the combination of the input approach of the device and input data which are used among said devices for an input according to a predetermined regulation While carrying out I/O media allocation for making the both sides of at least 1\*\* of the combination of the output method of a device and output data which use among said devices for an output according to a predetermined regulation correspond, choosing, and assigning It is multi-modal dialogue equipment characterized by providing the media allocation means which processes modification processing directions of I/O media allocation in order to make the change at the time of a carrier beam, and carries out new media allocation.

[Claim 9] While recording the I/O media allocation which has the media allocation storage slot which records I/O media allocation in said media

allocation means, and is chosen as said media allocation storage slot now Directions of I/O media allocation modification processing at the time of a carrier beam Claims 2 or 3 characterized by having the function to perform new media allocation using the I/O media allocation information currently recorded on said media allocation storage slot, 4 or 7, or multi-modal dialogue equipment given in 8 any 1 terms.

[Claim 10] It has the media allocation history storage slot which memorizes the sequence of the I/O media allocation chosen in the past in said media allocation means. When the sequence of the I/O media allocation chosen by then is memorized into said media allocation history storage slot and I/O media allocation is changed into it While adding the I/O media allocation then chosen to said media allocation history storage slot, directions of I/O media allocation modification processing at the time of a carrier beam Claims 2 or 3 characterized by having the function to perform new media allocation using the I/O media allocation information on the past currently recorded on said media allocation history storage slot, 4 or 7, or multi-modal dialogue equipment given in 8 any 1 terms.

[Claim 11] In said dialogue management tool, when said problem-solving equipment needs to acquire information required for a \*\* sake from a user to the problem solving of a system As opposed to the input request information which problem-solving equipment was generated and was sent through means of communications to the dialogue management tool the input request receptionist function for which said input means is asked so that I may have a user input said input required information -- or Information required for a dialogue management tool since the output to a user is constituted according to the I/O media allocation to which said output means is set the output request function for which it asks so that the output request information which asks problem-solving equipment for an output may be generated, it may send through said means of communications and problem-solving equipment may send said output required information through means of communications, and \*\*, while having at least one function inside It is multi-modal dialogue equipment according to claim 8 characterized by equipping said media allocation means with the function to perform new media allocation for directions of I/O media allocation modification processing using at least one of said input request information or said output request information at the time of a carrier beam.

[Claim 12] It is a deed about the input step which the input from at least one device for an input interprets [ step ] according to I/O media

allocation among the various devices for an input with which it differs for acquiring the information from a user in the dialogue approach which processes business by the dialogue with a user and outputs a result, and generates that information corresponding to an interpretation, and the processing according to the necessary object to this information response that generated. The output step which outputs the result of having been based on the output of the problem-solving step which outputs a result, and said problem-solving step, from the device for an output corresponding to I/O media allocation among various different devices for an output, The dialogue management step which directs modification processing of I/O media allocation according to the processing result from said problem-solving step, and manages the I/O by said input step and output step, In order to obtain the input from the input data given to at least one input device among said two or more input devices to said problem-solving step At least one of the combination of the input which chooses the device used among said devices for an input, and is generated for every input media In order to generate the expression shown to a user from the output of said problem-solving means And the inside of said device for an output, While choosing said I/O media allocation which is the combination of the both sides of at least 1\*\* of the combination of the output method of a device and output data which use according to a predetermined regulation and carrying out this input media allocation It is multi-modal dialogue equipment characterized by having the media allocation step which processes the modification in order to direct modification processing of these I/O media allocation, and generates new media allocation at the time of a carrier beam.

---

[Translation done.]

\* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
  - 2.\*\*\*\* shows the word which can not be translated.
  3. In the drawings, any words are not translated.
-

## DETAILED DESCRIPTION

---

### [Detailed Description of the Invention]

#### [0001]

[Field of the Invention] This invention relates to the human machine dialogue equipment which combines and has two or more available I/O means, and its control approach in an information processor.

#### [0002]

[Description of the Prior Art] A computer in recent years A keyboard, a mouse, a microphone, a touch panel, Two or more input means by which the classes of an image scanner, a video camera, a pen, a data glove, personal sensor, etc. differ, Namely, an input device can be equipped and various information can be inputted now. Various information also to output devices, such as a display and a loudspeaker, moreover, a spoken language and a sound effect, The function which can be outputted with various gestalten of music is equipped, these various I/O devices are used effectively, it depends for them, and the demand of a human machine interface which is easy to use is increasing. Moreover, using these I/O devices, while a user directs the object on a display with a finger, a command is uttered with voice in recent years, Equip the input means in which the complicated input which combined two or more input devices is possible, i.e., input media, and two or more output devices, such as a display and a loudspeaker, are used. For example, the expression of human being's face displayed on the display while telling the message in the spoken language from the loudspeaker, By equipping the output means in which the complicated output which combined an output device, and two or more of its contents of an output and formats with condition of telling nuance by the sound effect from a loudspeaker is possible, i.e., output media Development of the multi-modal interface technique in which the ease of using of an interface, the quality of signal transduction, and effectiveness are raised is performed briskly.

[0003] By the way, in order to realize transfer of the information between smooth and conventionally natural users and the application software on a computer, it considered that transfer of this information was a dialogue between a user and application, the dialogue plan was developed based on the dialogue regulation recorded beforehand, and it was carrying out realizing a multi-modal dialogue in the combination of the I/O approach decided in the dialogue regulation etc.

#### [0004]

[Problem(s) to be Solved by the Invention] However, the combination of the utilization method of the device and device which are used for I/O in

the phase of each dialogue, namely, by the method beforehand described in the dialogue regulation, media allocation (media location) For example, a situation for which it is not appropriate that circumference noise sound volume uses voice as an I/O means greatly cannot be coped with, Media allocation is fixed with the flow of a dialogue, the combination of an I/O means was chosen according to the situation, and there was a problem that it could not converse with a user.

[0005] For example, the case where the dialogue equipment which can be inputted by the user with voice is used is considered. In this case, with a current technique, since the technique of speech recognition has not matured enough, even if some users generate the same word repeatedly, that recognition goes wrong may occur. Generating many

this, when a peculiarity is in pronunciation too much, the probability of a recognition success by the dialogue equipment side becomes very low.

[0006] Nevertheless, conventional multi-modal dialogue equipment had the problem of the futility of time amount having arisen or making a user sensing stress in order to require such a user repeatedly to recarry out an input with voice.

[0007] Moreover, although having considered the itinerary system by voice directions of the configuration of searching the tourist resort of relevance from a database and outputting a retrieval result to an output means by speech recognition, for example to a name of a place input with voice this itinerary system asked the user for the name of a place input with voice A user On the other hand, since [ for example, ] the name of a place of "Ryoan-ji" (\*\*\*\*\*\*) was changed between "\*\*\*\*\*" and reading and was uttered, Since it did not have a means to perform I/O media allocation dynamically in conventional multi-modal dialogue equipment when a system fails in retrieval, it was difficult to recover communication with a user.

[0008] That is, although that a user does not know reading of the right proper noun "\*\*\*\*\*" has led to failure of communique SHOKU with a user If this is also changed to the format which makes input selection with the kanji, although possibility that communication recovery can be performed will be size very much Since it was not able to say that such I/O media allocation was changed according to a situation, it will lapse into a stiffness condition and it will be said that advice service cannot be performed as a result.

[0009] Moreover, the output gestalt is being fixed to the gestalt programmed beforehand when showing a user the content of the database retrieval result in a certain dialogue equipment, for example

from the former. That is, although the same output gestalt [ be / no relation in the output number of cases of a retrieval result ] is taken, for example, there is a retrieval result of dozens of affairs or hundreds of affairs, there were quite a few cases where it became the I/O method which is hard to use unintelligibly for a user of being as displaying the retrieval result which has only several affairs conversely by the tabular format \*\*\*\* [, and ]. [ reading them out with voice one by one ]

[0010] Furthermore, since a system was not able to perform a dialogue with a user to compensate for the media allocation which a user demands, the user of a system also had the problem that a system had to output and input according to the media allocation opted for and specified.

[0011] As for the I/O means which is easy to use for condition that another user likes carrying out the keyboard entry of the run string, by the user, differing is common, although it is not necessarily carrying out [ everybody ], for example, as for the input means which is easy to use for a user although there are various things in an input means, a certain user likes an input with voice.

[0012] However, the difference among these individuals was not taken into consideration in conventional dialogue equipment, but the combination of a fixed I/O means was offered, an I/O means cannot be selectively used now by liking of a user, and there was a problem that user-friendliness could not say it as a good thing.

[0013] Moreover, in conventional multi-modal dialogue equipment with two or more input means, there was a problem of it having been difficult which input means to be used or for a user to recognize \*\* whether it is available when there is which input means, and giving a user puzzlement and stress on the occasion of an input.

[0014] Furthermore, there was a problem that where of using voice for an input or an output even when it is not desirable to use voice -- the external noise becomes intense -- etc. cannot change the I/O method according to change of an external-environment-element in the conventional dialogue equipment which cannot perform media allocation (media allocation) dynamically, for example in the dialogue equipment in which a conventional input or a conventional output with voice is possible.

[0015] Thus, the user made it force with conventional multi-modal dialogue equipment to perform a system and a dialogue according to the combination of the I/O mode beforehand decided under the dialogue regulation which a system has, since it was not the interface which is

easy to use intelligibly for a user, on the occasion of the informational input, stress was sensed, and it is harder for a user to understand an output and, occasionally he had become also with the cause of failure in I/O.

[0016] Moreover, in voice word recognition mode, as it said that there might be the input, the failure, and the incorrect input resulting from recognition failure, the object of a dialogue had been un-attaining as a result also in the case of failure [ in the I/O resulting from the property of the specific selected I/O means ].

[0017] Then, the place made into the object of this invention is to offer the multi-modal dialogue equipment which can serve as an input/output interface which a user tends to use. It is in creating the multi-modal dialogue equipment which can offer the input/output interface which is easy to use for a user by having the function which chooses dynamically media allocation of I/O especially suitable at a certain event according to various situations as a front end of an application program in this invention.

[0018]

[Means for Solving the Problem] The multi-modal dialogue equipment concerning this invention A means to generate dynamically the approach of combining the approaches and those data for choosing and interpreting input data based on the rule defined beforehand in order to build the input data from the input data given to two or more input devices to [ from a user ] an application program, A means to generate dynamically the approach of choosing output data in order to generate the expression shown to a user from the output data of an application program, and the approach of the combination and an expression of those data based on the rule defined beforehand is provided. It is characterized by the ability to choose the combination of a means suitable about each of I/O at each time of a dialogue with a user dynamically, and continue a dialogue.

[0019] In case a user performs a computing system and a dialogue according to this invention, it becomes possible to choose and use the method of I/O using the I/O means which chooses the I/O method reflecting an available I/O means at each [ of a dialogue ] event, and a user tends to use for the user.

[0020] Moreover, according to this invention, media allocation (media allocation) is fixed with the flow of a dialogue, chooses the combination of an I/O means according to a situation, and becomes possible [ solving the problem of the conventional multi-modal dialogue equipment that it

cannot converse with a user ].

[0021] Furthermore, according to this invention, if it is conventional multi-modal dialogue equipment, a specific user will be repeatedly required to recarry out an input with voice, [ when there is a problem of the futility of time amount arising or making a user sensing stress ] It becomes possible to use a computing system, without solving this problem and a user sensing the futility and stress of time amount by providing the means which chooses and changes an I/O means dynamically.

[0022] Furthermore, according to this invention, when interpretation processing of an input goes wrong, by providing the means which chooses and changes an I/O means dynamically to having been difficult to recover communication with a user in conventional multi-modal dialogue equipment, this problem can be solved and handicap NYUKESHON with a user can be recovered promptly.

[0023] Furthermore, if the gestalt of a voice output is adopted in order to tell few retrieval results when showing a user the content of the database retrieval result in a certain dialogue equipment, for example from the former A retrieval result reads out dozens of outputs of the retrieval result with voice, even if there are hundreds of affairs, In order to take the same output gestalt [ be / no relation in the output number of cases ] and to make easy treatment of the retrieval result of a large quantity at reverse, when the tabular format is adopted, even if it is as a result of [ only with several affairs ] retrieval, it is [ fixed ] as displaying by the tabular format \*\*\*\*.

[0024] However, by establishing the means which chooses and changes an I/O means dynamically in this invention, and choosing and changing an I/O means dynamically, when becoming such an I/O method it is hard for a user to use unintelligibly, this problem can be solved and an I/O method intelligible for a user can be taken now.

[0025] Furthermore, according to this invention, the difference in the individual about the I/O means which is easy to use in conventional dialogue equipment is not taken into consideration, but it becomes possible to use the combination of the I/O approach which each user tends to use for each to the problem that the combination of a fixed I/O means was chosen.

[0026] Furthermore, according to this invention, it sets to conventional multi-modal dialogue equipment with two or more input means. As opposed to the problem of it being difficult for a user to recognize whether it being available when there are which input means being used

and which input means, and giving a user puzzlement and stress on the occasion of an input. By providing a means to display or put up the means which can be used for an input by the user, this problem is solved, a user is enabled to recognize an input means easily, and it becomes possible to offer the interface which is easy to use intelligibly for a user.

[0027] In this way, although a user is dependent on the quality of the information which a user and an application program search for, an amount, and an I/O means with the available computing system to be used, it becomes possible to use the human machine interface independent of an application program.

[0028]

[Embodiment of the Invention] This invention is providing a user with the I/O approach according to a situation by providing a means choosing the combination of a means dynamically, about each of I/O. By providing a means to realize the interface which is easy to use and to make a user know an available input means. A user is provided with a more intelligible input environment, the combination of an I/O means is changed further, failure in a dialogue with a user is recovered, the multi-modal dialogue equipment which can continue a dialogue is offered, and the detail of this invention is hereafter explained with reference to a drawing.

[0029] Drawing 1 is the block diagram of the multi-modal dialogue equipment concerning the example of this invention. As shown in drawing 1, the multi-modal dialogue equipment 10 of this invention consists of a total of four components, the input section 11, the media assignment-decision section 12, the dialogue Management Department 13, and the output section 14, performs interactive processing with a user through the physical device connected to a system, and performs intermediation of the problem-solving means (background application) 15 with a user.

[0030] The input section 11 is for incorporating the output of various kinds of input device IDs Va, such as a mouse device, a microphone, a television camera, a keyboard, a tablet, an image scanner, and a sensor, - IDVn, and the output section 14 generates the actuation output to various kinds of output device ODVa-ODVn, such as a display, a loudspeaker, a lamp, a printer, and a plotter.

[0031] The media assignment-decision section 12 is equipment for determining the combination of the utilization method of the various devices for the output for an input, and these various devices, i.e., media allocation, (media allocation), and has the function which can choose the

best combination which sometimes comes out.

[0032] Moreover, the dialogue Management Department 13 controls the approach of the dialogue between a user and a system, such as requiring media allocation or generating the plan of a dialogue based on the information delivered and received between the problem-solving means 15, the status information from the input section 11 and the output section 14, etc., and flow.

[0033] The problem-solving means 15 is the application which is the application which performs the reply and answerback to an input, and this issues the reply of the object by the dialogue, or does work of the object, for example, is good anything. [ of a sightseeing advice system, a commuter pass ticket sale machine, etc. ] A block diagram shows the detail configuration of the multi-modal dialogue equipment 10 of this invention to drawing 15 .

[0034] (Detail of the input section 11 and the output section 14) As shown in drawing 15 , the input section 11 of the multi-modal dialogue equipment 10 of this invention consists of three modules, the input device processing section 111, the input mode processing section 112, and the input media processing section 113, and the output section 14 consists of three modules, the output media processing section 141, the output mode processing section 142, and the output device processing section 143.

[0035] Here, with the mode, the input (the same is said of output) from one device should be interpreted and abstracted. For example, two or more interpretations and utilization by the mode called directions for actuation and issuance of a command with the mode components which interpret it are possible for the data inputted from the mouse device.

[0036] similarly, media have pointed out what can be passed to a referent by making into the content of directions the processing result which should combine the interpretation result from the mode of plurality (even -- also coming out -- good), and is obtained combining an input.

[0037] For example, mouse object directions--voice command media The input from the mouse device which are pointing devices, such as a personal computer, The voice command mode which can interpret a command with voice collectively and can interpret the voice which carries out command issuance and directs the command to application, The interpretation result in the mouse object directions mode in which the input from a mouse device is interpreted as actuation of directing the object for command actuation is considered as an input. They are

the media which perform telling processing "the command directed with voice to the object directed with the mouse is published" combining those inputs to application.

[0038] The input device processing section 111 which is the component of the <input section 11> input section 11 considers the data transmitted from each physical input device as an input, and has in a component the input device processing components which perform the minimum processing for interpreting the data, such as conversion to digital data. Device processing components are prepared for every input device, and operate to independent and juxtaposition. After each device processing component performs the above-mentioned processing to the input to a corresponding device, it is sent to one or more corresponding input mode processing sections 112. For example, microphone device processing components incorporate the voice which the user inputted from the microphone with a background noise, for example, change it into a digital audio signal with a sample frequency [ of 12kHz ], and a quantifying bit number of 16 bits, and output the signal to the input mode processing section 112.

[0039] The input mode processing section 112 has the components which process the input to each abstract input mode in a component. One input mode processing component corresponds to one abstract input mode, and abstracts and interprets the data which were sent from the one input device processing section and which were operated orthopedically for every abstract mode. Each input mode component has the response relation of 1 to 1 or one-pair \*\*, as shown in input device components and a table 1.

[0040] The result interpreted in each input mode processing section 112 is sent to the input media processing section 113. For example, voice command input mode processing components search for feature-parameter time series by the frequency analysis according the sent digital audio signal to a fast Fourier transform, perform speech recognition based on methods, such as a compound similarity method and a hidden Markov model, further, for example, obtain the highest vocabulary of likelihood out of the given command vocabulary for recognition, and output the information to the input media processing section 113.

[0041] The input media processing section 113 generates the input integrated by combining the interpretation result for every mode of the input of a user obtained from the input mode processing section 112 per each abstract media, it chooses the input from required media in from

among two or more obtained media inputs, following directions of the media assignment-decision section 12, and gives the input command and the data to the problem-solving means 15.

[0042] The component of the input media processing section 113 is the media selecting switch SW which specifies which is sent to the problem-solving means 15 among the processing results of those components as input media processing components 113a prepared for every abstract media. Each input media processing components 113a receives the output data (interpretation result in abstract mode) from one or more input mode components as an input, unifies them, and sends the interpretation result as media to the media selecting switch SW.

[0043] The media selecting switch SW chooses and sends the media output sent to background application according to directions of the dialogue Management Department 13. For example, mouse object directions-voice command media input-process components consider the output of mouse object directions mode input-process components, and the output of voice command mode input-process components as an input, and generate and output the interpretation as media combining them. And for example, when mouse object directions-voice command media are chosen as current input media, the media selecting switch SW hands over only the interpretation result of the input by these media for the problem-solving means 15.

[0044] The output media processing section 143 which is the component of the <output section 14> output section 14 chooses the output data from the problem-solving means 15, following the allocation information on the media which the dialogue Management Department 13 gives (allocation information), they distribute to each abstract output media, and it processes them for every media, and delivers the data which divide and unify them per output mode and the output mode deals with on each mode processing components of the output mode processing section. In order to perform such processing, the output media processing section 143 consists of three modules, output media distribution switch 143a, output media processing section 143b, and media mode interface 143c. Following directions of the dialogue Management Department 13 in the output data which background application has sent, media distribution switch 143a chooses the data outputted actually (to user), and distributes it to abstract media processing components.

[0045] The output media processing components 143 have the output media processing components corresponding to each abstract output

media as a component, and process and send out the output data which were able to be distributed to the media in their duty by media distribution switch 143a to the form which can be outputted to the media. Media mode interface 143c has the media mode interface components corresponding to each abstract output mode 1 to 1 in a component. Each media mode interface component unifies the output from one or more abstract media processing components which the mode takes charge of, and delivers data to corresponding output mode processing components.

[0046] The output mode processing section 142 has in a component mode device interface 142b which unifies output mode processing components 142a and the output data for every mode which process the output to each abstract output mode for every output device. One output mode processing component corresponds to one abstract output mode, processes every output mode to the media data sent from the media mode interface components corresponding to the mode, and delivers it to mode device interface 142b.

[0047] One component of mode device interface 142b is equivalent to one output device component, the output from one or more abstract device processing components which the device takes charge of is unified, and data are delivered to corresponding output device processing components. The output device processing section 141 has as a component the output device processing components prepared for every output device. Each output device processing component drives the output device which takes charge of the output data of a corresponding media device interface as an input command and data.

[0048] (The detail of the media assignment-decision section 12) As shown in drawing 15, the media assignment-decision section 12 of the multi-modal dialogue equipment 10 of this invention consists of three storage components of the mode-media allocation conversion table attaching and showing the present media allocation storage slot and the sign 124 which attach and show the media [ degree ] assignment-preference table and the \*\*\* sign 123 which attach and indicate a sign 122 to be a module called the media allocation generation section, as a sign 121 was attached and shown. Among these, it combines (media allocation), and the media allocation generation section 121 generates the suitable combination and expression approach of *Perilla frutescens* (L.) Britton var. *crispa* (Thunb.) Decne. as the data actually inputted into application, and its data actually given to a user with reference to the media [ degree ] assignment preference table 112 at the question of the

dialogue Management Department 13 according to doubling, and transmits it to the dialogue Management Department 13.

[0049] As shown in drawing 15 , (Detail of the dialogue Management Department 13) The dialogue Management Department 13 of the multi-modal dialogue equipment 10 of this invention The condition of the I/O mode obtained from the input media processing section 113 and the output media processing section 143, The condition of the data exchange with the problem-solving means (background application) 15 The plan of the dialogue [ by the description information common to two or more users for the input/output request from advice or the problem-solving means 15 of the condition of the problem-solving means 15 to the problem-solving means 15 and the information about the control approach of the problem-solving means 15 ] based on reception and those information as an event is generated.

[0050] And the approach and flow of the dialogue between a user and a system are controlled by directing for the input media processing section 113, the output media processing section 143, and the problem-solving means 15 in the activation process of the plan.

[0051] If it considers that transfer of the information between other modules, such as the input media processing section 113, the output media processing section 143, the problem-solving means (background application) 15, and the media assignment-decision section 12, is a dialogue as a framework which performs such processing, for example The approach using the state-transition model like the voice dialog system indicated to JP,5-216618,A, Sacerdoti, E.D. work "A Structure for Plans and Behavior", American It is realizable by using a hierarchy plan composite system which will exist in Elservier \*\* and 1977.

[0052] The dialogue Management Department 13 has the module called a dialogue planner, as the sign 131 was attached and shown, although it does not illustrate this dialogue planner 131 with two storage tables, the dialogue regulation storage dialogue primitive storage table 132 and 133, it has three storage components called the plan storage section, an initial target record slot, and a current target slot, and this builds and performs a dialogue plan in it.

[0053] The above is the configuration and its function of this equipment. The flow chart of actuation of the dialogue Management Department 13 which is the nucleus module of this equipment is shown in drawing 19 . Based on this drawing, the usage of the principle of operation of this equipment and data is described.

[0054] First, if this equipment is started, the dialogue planner 131 of the

dialogue Management Department 13 will develop all dialogue plans in order in the plant storage section which is not illustrated in the dialogue Management Department 13 based on ejection and its target from the initial target record slot to which the initial target for plan generation is set beforehand.

[0055] One item of the plan storage section consists of five slots, "Node ID", a "target", a "condition", a "subnode ID list", and a "result", and calls this a plan node. If one certain target is set up, one plan node will be generated, the set-up target will be substituted for a target slot, and a unique notation will be assigned to a node ID slot at each plan node.

[0056] One of five of "a success" unsettledness and "during processing", "failure", and "retry" is substituted for a condition slot. ID of the plan node corresponding to the sub target developed from the content of the target slot is contained in the form of a list by the subnode ID list slot.

[0057] The activation result of the target of a plan node for the slot to belong is contained by the result slot. Henceforth, for example, the thing of a "plan [ under which the value of a node ID slot is p000 ] node" is only described to be "a plan p000" or the "subplan p000."

[0058] In order to develop a certain dialogue target at the sub target, the dialogue Management Department 13 has the dialogue regulation storage table 132. Two or more items showing a dialogue regulation as shown in the dialogue regulation storage table 132 at drawing 2 are stored, and the each is called a dialogue regulation node. One dialogue regulation node consists of four slots, "Rule ID", a "target", a "usage condition", and a "sub target list." As for "Rule ID", a unique notation is assigned to each set talk regulation like the node ID of a plan node.

[0059] The target which the regulation attains in each set talk regulation node is dedicated to a target slot, the sequence of the sub target (generally plurality) for a target to be developed and generated is dedicated to a sub target list slot, and the conditions for applying a regulation are dedicated to a usage condition slot, respectively.

[0060] The dialogue planner 131 searches for the dialogue regulation stored in the dialogue regulation storage table 132 sequentially from a top, and applies the first dialogue regulation which suits conditions. Generation and activation of a dialogue plan are realized by repeating recursively actuation of searching for and applying the regulation with which are satisfied of the given target out of a dialogue regulation, and developing and performing it at a sub target.

[0061] And when an initial target is satisfied eventually, the whole

dialogue is successful and this equipment stops. Henceforth, for example, the value of a rule ID slot only describes the thing of "the dialogue regulation node which is r00" to be the "dialogue regulation r00" or "a rule r00."

[0062] The dialogue planner 131 dedicates and manages the target of the dialogue in a certain event into a current target slot. Before searching for the dialogue regulation which develops the target of the present target slot, the dialogue planner 131 looks for the dialogue primitive storage table 133 which contained the dialogue primitive which is the smallest unit of the dialogue in which immediate execution is possible as shown in drawing 3 .

[0063] A dialogue primitive node, a call, and this consist one item of the dialogue primitive storage table 133 of three slots, "FAKUTO ID", a "target", and a "specification." When the content of the current target slot can simplify to the value of the target slot of a certain dialogue primitive node in the dialogue primitive storage table 133, dialogue PURANAN 131 performs the content which carries out the average of the target to a dialogue primitive, and is described by the corresponding specification slot of a dialogue primitive node.

[0064] The primitive of the dialogue of the inter module which the dialogue Management Department 13 manages and performs is either of the input/output request to the inquiry, the reply and the setting-out command to I/O media components and input/output command of suitable media assignment to the media assignment-decision section 12, and the problem-solving means 15.

[0065] The thing of the dialogue primitive node [ henceforth, ] of a "FAKUTO ID slot whose value is "f00"" is only described to be the "dialogue primitive f00" or "FAKUTO f00."

[0066] When there are both a dialogue primitive which can be simplified at the target in the event of being at the time of expansion and activation of a dialogue plant, and a dialogue regulation, the target goes wrong. When a certain target goes wrong, the dialogue planner 131 is the approach same with backtracking a calculating machine in the activation process of the "PROLOG language" which is one of the computer program language, and performs back track of the expansion process of a dialogue plant, and retry of a dialogue regulation.

[0067] Namely, a certain dialogue target Pi Activation is faced and it is the dialogue regulation Ri. When it applied and fails, it is dialogue planner 131Pi. It searches for other applicable dialogue regulations or a dialogue primitive, if there is an applicable thing, the dialogue regulation or a

dialogue primitive will be applied, and it is Pi. The target to be satisfied is retried.

[0068] Pi the case where applicable all dialogue regulations and dialogue primitives go wrong -- the dialogue planner 131 -- dialogue target Pi Dialogue target Pc which was successful immediately before up to -- retracing one's steps -- the same -- the dialogue target Pc It retries. As long as failure of target execution continues, when back track of a target is generated continuously, their steps are temporarily retraced to an initial target and the initial target itself goes wrong, it means that the whole dialogue had gone wrong and this equipment stops.

[0069] The media assignment-decision section 12 is called from the dialogue planner 131 in the form of activation of the dialogue primitive "MA inquiry" which is an inquiry of multimedia allocation (MA). The media assignment-decision section 12 opts for suitable media allocation using the media [ degree ] assignment preference table 122 as indicated to be the condition of the media allocation in the event of the inquiry contained by the present media allocation storage slot occurring to drawing 6 at the event.

[0070] one item of the media [ degree ] assignment preference table 122 -- a media [ degree ] assignment preference node and a call -- each -- a media [ degree ] assignment preference node has four slots, "Node ID", the "present media allocation", "conditions", and "media [ degree ] allocation."

[0071] The media allocation generation section 121 searches for the media [ degree ] assignment preference node which the conditions at the time of the content and the media assignment-decision section 12 of the present media allocation storage slot being called suit out of the media [ degree ] assignment preference table 122, and generates the candidate of media [ degree ] allocation.

[0072] Next, the media allocation generation section 121 searches for the mode-media allocation conversion table 124 as shown in drawing 7 from the content of the media [ degree ] allocation slot of the discovered media [ degree ] assignment preference node, and verifies the constraint at the time of applying a certain media allocation. The node corresponding to mode-MEDEAA allocation, and a call and each node corresponding to mode-media allocation consist one item of the mode-media allocation conversion table 124 of four slots, "ID", a media allocation name", "constraint", and an "I/O mode group."

[0073] In case the media allocation generation section 121 performs constraint verification of media allocation, it searches for the node

corresponding to the mode-media allocation which agrees with the media allocation name for which the slot of a media allocation name searches, and it verifies [ that it is / constraint / at the time of performing actually media allocation described by the constraint slot of the node corresponding to mode-media allocation / condition / in an ejection and inquiry event / satisfied in the constraint, and ].

[0074] If this constraint is satisfied, that media allocation candidate will be eventually chosen as media [ degree ] allocation, and the dialogue planner 131 will reply to him. In case the decision of media allocation is requested from the media allocation generation section 121 and it looks for a media [ degree ] assignment preference table here, when the value of the present media allocation slot of a certain media allocation node or a condition slot is “-”, the value of the slot is not used for verification.

[0075] For example, the media [ degree ] allocation node whose content of the present media allocation slot is “-” can be adapted, if the content of the condition slot and the content of the constraint slot of a mode-media conversion table conform whatever the value of the present media allocation storage slot may be. Moreover, into the present media allocation slot and media [ degree ] allocation slot of a media [ degree ] allocation node, it can describe also in the form of the group of input media and output media in addition to the content of media allocation.

[0076] Moreover, it is possible to use a variable for these input media or output media. For example, although the present media allocation slot value is “(spoken language, XXX)” and the value of a media [ degree ] allocation slot has become “(pen character string, XXX)” at drawing 6 in the media [ degree ] assignment preference node whose value of a node ID slot is “m06” A media [ degree ] assignment preference node is applicable if it is a “spoken language”, the input media, i.e., the input media by which current setting out is carried out, of the media allocation which is the value of the present media allocation storage slot, and other constraint [ whatever the output media ] is satisfied.

[0077] At this time, the value by the side of the output media of the present media allocation storage slot is temporarily assigned to the value by the side of the output media of the present media allocation slot, i.e., variable”, “XXX”, and that substitution result is reflected in the media [ degree ] assignment preference node to which that present media allocation slot belongs.

[0078] And retrieval and verification of constraint of a media [ degree ] allocation slot are performed by matching with the value of the I/O media slot of a mode-media allocation conversion table. For example,

when the value of the present media allocation slot is "human", call conditions presuppose that the media allocation generation section 121 was called by "input failure."

[0079] In order that the media allocation generation section 121 may opt for media [ degree ] allocation, it looks for the media [ degree ] assignment preference table 122 first, but among the media [ degree ] assignment preference nodes in drawing 6, since, as for the media [ degree ] assignment preference node of "m01" to "m05", conditions do not suit, the value of a node ID slot is not chosen.

[0080] Next, although the value of a node ID slot tends to verify whether the media [ degree ] assignment preference node of "m06" can be adapted, since the value of the present media allocation slot of this media [ degree ] assignment preference node is described by in the form of the group of input media and output media, the media allocation generation section 121 reconfigures the value of the present media allocation storage slot in the group of input media and output media first.

[0081] therefore, the media allocation generation section 121 searches for drawing 7, and the value of the media allocation slot of the node corresponding to mode [ in which the value of a node ID slot is "mm01" ]-media allocation from the same "human" thing as the value of the present media allocation storage slot the media allocation from the value of the I/O mode slot of this node corresponding to mode- media allocation -- "human" input media -- a "spoken language" mode simple substance -- it is -- output media -- "a spoken language" -- it is the combination in the mode and "face image" mode -- "(a spoken language --) It acquires that it is the I/O media group spoken language + face image."

[0082] The media allocation generation section 121 is the value of the present media allocation slot of the media [ degree ] assignment preference node of the affair in which the value of Node ID is m06 using this information. "(spoken language, XXX)" and simplification are attained and a "spoken language + face image" is temporarily substituted for Variable XXX.

[0083] This substitution is reflected in the media [ degree ] allocation slot of the same media [ degree ] assignment preference node, and the value of this slot becomes "(pen character string, spoken language + face image)." Next, the media allocation generation section 121 verifies the value "call condition = input failure" of the condition slot of the same media [ degree ] assignment preference node, and is successful in

UFR-AUG-9112 20/00  
this case. Furthermore, the media allocation generation section 121 searches for the mode-media allocation conversion table like drawing 7, and verifies constraint of the node corresponding to mode [ in which the value of an I/O mode slot is "(pen character string, spoken language + face image)" ]-media allocation.

[0084] When that constraint is satisfied temporarily, selection of media [ degree ] allocation is successful, and the media allocation called "the pen character string input and human response" which is the value of the media allocation name slot of this node corresponding to mode-media allocation is substituted for the present media allocation storage slot, and it is returned to the dialogue Management Department 13 as a media allocation result (media allocation result). Termination of verification of the media [ degree ] assignment preference node to which the variable belongs cancels temporary substitution to Variable XXX. The thing of the media [ degree ] assignment preference node [ henceforth, ] of a "node ID slot whose value is "m00"" is only described to be "the media [ degree ] assignment preference node m00." The thing of the node [ moreover, ] corresponding to mode [ in which a value is "mm00" ]-media allocation of "ID slot" is only described to be "the node mm00 corresponding to the Mohd Media allocation."

[0085] In order to make easy to understand [example of concrete actuation (first example of operation)] this invention, the example at the time of applying this equipment to the application of the hotel advice in a tourist resort as an example is described below. It is equivalent to the problem-solving means 15 which appeared by explanation of the application of the hotel advice in a tourist resort here, i.e., former.

[0086] The dialogue primitive shown in the dialogue regulation and drawing 3 which are shown in the dialogue regulation storage table 132 of the dialogue Management Department 13 at drawing 2 is memorized beforehand. Moreover, the initial target of a "interactive session" is given beforehand, and is memorized by the initial target storage slot.

[0087] When starting this equipment first, the dialogue planner 131 calls the regulation "r00" with which it is [ an initial target ] satisfied of ejection and its target from an initial target storage slot. And it is developed as the dialogue planner 131 shows a dialogue regulation "r00" to drawing 4.

[0088] The dialogue planner 131 calls the user appearance which is an early sub target, a dialogue regulation "r11" is developed, and a dialogue regulation "r12" is developed further. In order that the dialogue planner

131 may not have the dialogue regulation which can be adapted in the dialogue regulation storage table 132 in activation of a sub target "MA inquiry (standby, MA1)", the dialogue primitive storage table 133 shown in drawing 3 is searched.

[0089] Consequently, the dialogue planner 131 discovers a dialogue primitive "f01", performs "a media assignment-decision section call" in the item of a specification, calls it to the media assignment-decision section 12 at the event, and conditions are suited and it asks available media assignment.

[0090] The media assignment-decision section 12 opts for the media allocation which should lengthen the media [ degree ] assignment preference table 122 as shown in drawing 6 based on the media allocation at the event, and the situation of available media or others, and should be chosen as a degree. In this case, a table element "m01" is applied and media allocation of "standby" is chosen.

[0091] Next, the media assignment-decision section 12 checks the constraint at the time of applying the media allocation which lengthens the Mohd-media allocation conversion table of drawing 7, and corresponds. In this case, if the input sensor device is available, the media assignment-decision section 12 will choose media allocation of "standby" eventually, and will answer the dialogue Management Department.

[0092] The dialogue Management Department 13 notifies the media which should be used for the input media processing section 113 and the output media processing section 143, respectively based on the media allocation which the media assignment-decision section 12 has answered. Thereby, the input media processing section 113 and the output media processing section 143 lengthen the Mohd-media allocation conversion table according to the advice, and an input/output channel is set up by changing the condition of media selecting-switch SW and media distribution switch 143a, respectively.

[0093] In standby media allocation of this example, an aperture and an output channel output message waiting like drawing 20 only for a personal sensor to a screen as an input channel. By the above processing, a sub target "p11" tends to be successful and the dialogue planner 131 tends to put the following sub target "detection standby" into activation. This sub target is a dialogue primitive as it is memorized by the dialogue primitive storage table 133 of drawing 3, and it is performed immediately.

[0094] At this time, the condition of the plan storage section in the

dialogue Management Department 13 has become like drawing 5 . A dialogue primitive "user appearance" continues a waiting state until the condition of personal sensor appearance media changes.

[0095] If a user stands in front of equipment, an input sensor will react, a dialogue primitive "user appearance" is successful, a subplan "p01" is satisfied, and the following sub target "MAR selection (first stage)" is called. This sub target is further developed by "r12", and like the time of selection of standby media, in order to perform a sub target "MA inquiry", the dialogue planner 131 calls the media assignment-decision section 12, and asks the media allocation for dialogue initiation.

[0096] The media assignment-decision section 12 chooses the media allocation with which are satisfied of conditions and constraint by lengthening the media [ degree ] assignment preference table 122 and the Mohd-media allocation conversion table 124 which self holds like the case of standby media allocation.

[0097] Consequently, in the case of this example, media allocation of being "human" is chosen, and it is answered by the dialogue Management Department 13. The dialogue Management Department 13 is directing the media allocation to which it was answered in the input media processing section 113 and the output media processing section 143, and performs the dialogue based on media allocation "Hugh Mann."

[0098] this media allocation -- that it is "human" is the media allocation which can carry out a natural dialogue with voice to an input by using a face image and a spoken language for speech recognition and an output. Therefore, that a user should just perform a question, want and a reply, a response, etc. in his voice, an equipment side will display and combine a face image with a display, answerback, a reply, etc. to it will be performed using a spoken language, and its human being is intelligible and tends to treat them.

[0099] Furthermore, if data are given to the media assignment-decision section 12 from the input media processing section 113 or the output media processing section 143, for example as voice input Mohd or voice output MORT is unusable since a background noise is loud when above, media allocation "Hugh Mann" will not be chosen, but another media allocation will be performed, and a dialogue will be continued. For example, as not influenced at all of a background noise, it is a condition of the character representation according [ an output ] to a screen according [ an input ] to a keyboard.

[0100] It becomes possible in this way to provide a user with the interface which can choose suitable I/O media now dynamically and is

easier to use them by having considered as the configuration which adds a means to opt for the media allocation for the basic configuration of conventional multi-modal dialogue equipment with suitable this invention by this according to a situation according to the situation of a dialogue, and an I/O means available at the event.

[0101] Although the above is that a multi-modal dialogue equipment side chooses the optimal various I/O media dynamically corresponding to the occasional situation and the user was provided with the interface which is easy to use, it can also carry out as the configuration which offers the interface with which it is carrying out selection assignment, and each user tends to use for arbitration the thing for which a user side wants to use it of the various I/O media.

[0102] What is necessary is just to add a means to put up the media which a user can further input to the above-mentioned configuration in multi-modal dialogue equipment, in order to make at arbitration the thing for which a user side wants to use it of the various I/O media the configuration which carries out selection assignment. By considering as such a configuration, it can know easily with what kind of means a user can input, and it becomes possible to offer the interface which is easier to use.

[0103] For example, in media allocation "Hugh Mann" of the above-mentioned example of operation, by displaying the icon of a microphone on a screen like drawing 16, a user is made to know that an input with voice is possible, and the effectiveness of urging an input with voice can be expected. Then, in the process of being satisfied with a dialogue regulation "r03" of a sub target "an initiation session", session initiation demand of a user and recognition of a system are performed, and it becomes possible to operate "A" and "B" of a scenario as an example which are shown in drawing 8 in this way.

[0104] Namely, scenario A

Thereby, in an advice system, a sensor detects this, turns ON a display, displays the title screen of a system, and displays on a screen the prompt one traveler stands in front of an advice system and which stimulates an input with a sound effect.

[0105] Scenario B

A traveler says with voice, "Please look for a hotel", and a system answers to this "The system understood."

[0106] If the initiation session of such a content is successful, the dialogue Management Department 13 will announce that the task was started by the application which realizes the advice system which is a

problem-solving means. In this case, the application concerned which is a problem-solving means requires "acquisition of the initial retrieval conditions for database retrieval, i.e., the "date of arrival", a start day", "manpower", and a "budget" of the dialogue Management Department 13 according to that advice.

[0107] Based on this demand, the dialogue Management Department 13 demands the decision of the again suitable media allocation for the media assignment-decision section 12. By performing the \*\*\*\* dialogue shown in "C" of the scenario of drawing 8 using voice, displaying the condition of the retrieval conditions under input based on the answer like a screen as set media allocation "a variable display voice input/output" and shown in drawing 17 The dialogue Management Department 13 satisfies the demand of the application which is a problem-solving means.

[0108] (Second example of operation) Second example \*\*\*\*\* of operation which changed scene expansion next. By the dialogue to the scenario "C" of drawing 8 , the initial condition of database retrieval is acquired, and a problem-solving means (background application) performs retrieval processing, and requires the output of a retrieval result of the dialogue Management Department 13 after retrieval termination.

[0109] The outline of a retrieval result, i.e., the number of cases of a retrieval result and database elements in every affair, at this time, the problem-solving means 15 side with an output request Namely, a "name", "reading of a name", an "address", "magnitude", "the class of chamber and accommodation charges of each part store", In addition, it is announced that the "description", a "appearance photograph", a "map", "the way and time of concentration" to there, a "nearby station", and "a neighboring scenic spot" are outputted to the output media processing section 143 in the output section 14 of multi-modal dialogue equipment 10 from the problem-solving means 15 side.

[0110] At this time, the plan storage section of the dialogue planner 131 is in the condition of having succeeded to the sub target "p42" of drawing 10 . Although the dialogue planner 131 tends to search a dialogue regulation and it is going to apply "r07" first, in order to satisfy the "candidate decision" which is the following sub target "p43", since a usage condition does not suit, it fails, and other dialogue regulations are searched.

[0111] Next, the dialogue planner 131 chooses "r08" and "a selection demand" which is a sub target "p431" is developed. This sub target is

similarly developed by the sub target system train [p4311, p4312, p4313] under a dialogue regulation "r10."

[0112] Among these sub target system trains, a sub target "p4311" is developed further, the plan storage section will be in a condition like drawing 10, and 12 will be further called by the media assignment decision according to a dialogue regulation "r12." Although the same processing as old description is performed in the media assignment-decision section 12, since there are few retrieval results as three affairs, in retrieval of the media [ degree ] assignment preference table 122, a media [ degree ] assignment preference node "m3" is suited and chosen as conditions.

[0113] Consequently, "touch panel display selection and element name voice reading raising" are chosen as I/O media allocation for a user to choose a candidate from retrieval results, and the media allocation is performed by the dialogue Management Department 13.

[0114] Like drawing 18, about the accommodations of three affairs on a screen by this, respectively "A name", While the output media processing section 143 chooses the touch panel carbon button with which an "address", a "photograph", "magnitude", and "a neighboring name" have been arranged from the data poured from the problem-solving means 15 and sets it Voice will be outputted for the message to which the name of each housing is read out with voice and the selection is urged like the scenario "D" of drawing 8 from a loudspeaker.

[0115] In the case of a certain dialogue equipment, the output gestalt same irrespective of the output number of cases was taken from the former, for example, there were quite a few cases where it became the I/O method which is hard to use unintelligibly for a user of being as displaying the retrieval result which has only several affairs conversely by the tabular format \*\*\*\* [, and ]. [ reading out the retrieval result of dozens of affairs or hundreds of affairs with voice ]

[0116] For example, also when there were many retrieval results of a database, with conventional dialogue equipment, it had occurred that the name of all housings is read out like the dialogue of the scenario "D" of drawing 8.

[0117] This dialogue equipment has introduced the device in which the gestalt of I/O is dynamically set up according to a situation from the former in addition to the dialogue control mechanism based on a certain hierarchy plan generation method or a state-transition graph method that such a problem should be coped with. Therefore, when there was much object number of cases shown to a user, as according to this

equipment the dialogue Management Department 13 asked suitable media allocation to the media assignment-decision section 12 at the event, and it set up media allocation dynamically corresponding to the content of a reply, and it said that retrieval result presentation of a tabular format was chosen, for example, according to the occasional situation, retrieval result presentation can be carried out with the optimal gestalt.

[0118] For example, it sets in the dialogue of the scenario "D" of drawing 8. In order to make a user choose the thing of a wish from the retrieval result of the accommodations which the presentation object number of cases of a retrieval result is three affairs as mentioned above, and are these three affairs I/O media allocation in the form which the node "m04" of the media [ degree ] assignment preference table 122 was chosen in the media allocation generation section 121, consequently carried out to one affair like drawing 18, and cleft the area of a screen comparatively was made.

[0119] However, when the retrieval result number of cases is ten affairs temporarily also in the same scene, I/O media allocation which does not give a user stress in the form where only the retrieval result number of cases is tell with voice is make, without set to "m05" the media [ degree ] assignment preference node choose in the media allocation generation section 121, and the screen output of a tabular format like drawing 21 be assign by this, and read out an accommodations name with voice.

[0120] In this way, according to this equipment, corresponding to the informational quality and the informational amount which require an output, it is easy to accept application (problem-solving means) in a user, and it becomes possible [ setting up the combination of an intelligible output means ].

[0121] Furthermore, in order to give a big burden like [ in the case of the scenario "G" of drawing 8 ], for example in case a user chooses a result also with the output gestalt in a tabular format when there is very much number of cases of a database retrieval result, it is not appropriate to show a retrieval result according to the demand of application (problem-solving means) itself.

[0122] In this case, it is as follows. Now, like the case of the scenario "D" of above-mentioned drawing 8, retrieval processing by the problem-solving means is performed, and the plan storage section of the dialogue planner 131 presupposes that the dialogue is progressing to a sub target "p43111" like drawing 10.

[0123] Since there is no media [ degree ] assignment preference node 122 to which there are many retrieval result elements as 534 affairs, and conditions suit the media [ degree ] assignment preference table of drawing 6 in the case of this example, the sub target "p43111", i.e., "MA inquiry", to call the media assignment-decision section 12 will go wrong. In response to failure of "p43111", the dialogue planner 131 performs the same relapse processing as the PROLOG interpreter which is one of the well-known computer languages.

[0124] That is, it has a relapse into "MA selection (alternative presentation)" of "p4311" which is the high order gall of "p43111" first, and retries whether there is any dialogue regulation or dialogue primitive applicable to "p4311." in this case, the dialogue regulation "r12" which went wrong -- "a selection demand" of "p431" which "p4311" goes wrong and is the high order gall of "p4311" since there is also no unexpectedly applicable dialogue regulation and dialogue primitive -- their steps are still retraced.

[0125] Furthermore, since this "p431" goes wrong similarly, their steps are retraced to the "candidate decision" of "p43", it is similarly searched for a dialogue regulation or a dialogue primitive applicable for retry of this gall, the dialogue regulation "r09" of drawing 2 is applied, and a sub target sequence [p4301, p4302, p4303] is developed like drawing 11.

[0126] On the occasion of activation of "retrieval condition narrowing down" of "p4301", further, the dialogue planner 131 applies a dialogue regulation "r14", and generates a sub target sequence [p43011, p43012, p43013].

[0127] Next, the inquiry and problem-solving means 15 side reports a non-inputted variable list [the planned sightseeing site, a chamber format, and an accommodations format] for the variable which the dialogue Management Department 13 has not used for housing retrieval yet to the problem-solving means 15 to the request in the activation process of the "sheep input-variable inquiry" of the sub target "p43011" which is a dialogue primitive.

[0128] A variable "the planned sightseeing site" is taken out from from among non-input-variable lists by activation of "member (an input variable, [the planned sightseeing site, a chamber format, an accommodations format])" of the following sub target "p43012."

[0129] And in order to make a user input it, using the variable as additional conditions, a dialogue regulation "r17" is applied to "the additional condition input" of the following sub target "p43013", and a

sub target sequence [p430131, p430132, p430133] is generated. [0130] In the activation process of these sub target sequences, a dialogue like the scenario "G" of drawing 8 is performed, and this equipment becomes possible [ recovering failure in media allocation and continuing a dialogue with a user ] in this way.

[0131] Drawing 11 shows the condition of the plan storage section when the utterance S14 of the system of the scenario "G" of drawing 11 drawing 8 is emitted.

[0132] (Third example of operation) The third example of operation which changed scene expansion next is described. Suppose that the user found a route and time of concentration like the scenario "E" of drawing 8 after having a dialog to the scenario "D" of drawing 8. That is, since it did not have a means to perform I/O media allocation dynamically in conventional multi-modal dialogue equipment when a user asks for the output in different media allocation from the media allocation from the media allocation which dialogue equipment 10 opted for and outputted, it was impossible to have met the demand of such a user primarily. However, such a situation can be coped with with the equipment of this invention.

[0133] When such a situation is encountered, this equipment can separate control of a dialogue, and assignment of I/O media, and can make media assignment preference of the flow of a dialogue of the media assignment-decision section 12 independently. Therefore, management in such a situation is enabled.

[0134] When there was a demand of a user like U6 in the scenario "E" of drawing 8, the condition of the plan storage section of the dialogue planner 131 was in the condition of waiting for a user's input in the condition under processing of the "own alternative" of a sub target "p4313", like drawing 12. Since a user's input was not the selection courtesy of a candidate but a re-output request in accordance with media allocation modification and it here, a sub target "p4313" goes wrong.

[0135] Relapse processing occurs like the case of the second example of operation with this failure, "a selection demand" of "p431" which is the parent target of "p4313" goes wrong, and the dialogue regulation "r15" of drawing 2 is retried. The dialogue planners 131 are the conditions of "a user demand (output [a route, time of concentration])" in the activation process of this dialogue regulation, and selection of the suitable media allocation for the media assignment-decision section 12 is requested. By this request, the media assignment-decision section 12

chooses the media allocation "a name, a photograph, and map touch panel display selection and route voice reading raising." And this selected media allocation is passed to the dialogue planner 131.

[0136] While the dialogue planner 131 changes the output media allocation at the time of displaying the alternative of accommodations according to this selection, displaying that "name", a "photograph", and a "map" on a display about each accommodations of three affairs and setting up a touch panel carbon button, utterance [\*\*\*\*\* like S11 in the scenario "E" of drawing 8 is by bus as about 1 hour and a half. By train, from the place which went about 30 minutes, it changes to a bus and the countrified hermitage takes about 30 minutes. The Okuyama \*\* is in the place along which he walked about 30 minutes, after riding about 30 minutes by bus from here. The content " is uttered, a user is provided with information, and selection of the candidate of accommodations is again required of a user. The condition of the plan storage section in this event is shown in drawing 13 .

[0137] In this way, according to this equipment, using the I/O approach which a user demands, the continuation of a dialogue in the form where the user had the leadership according to a user's interruption is attained, and it becomes possible to offer the interface which is easy to use for each user.

[0138] (Fourth example of operation) the example of management actuation in a case since the content of the voice input obtained from the user was not right, as the system failed in speech recognition processing -- next, it states as fourth example of operation. scenario "H of drawing 8 -- " -- since it did not have a means to perform I/O media allocation dynamically in conventional multi-modal dialogue equipment, like when a system fails in speech recognition processing, since the user changed the name of a place between readings although the system asked for the voice input from a user, it was difficult to recover handicap NYUKESHON with a user.

[0139] However, if it searches for other dialogue regulations of being able to be satisfied with this equipment of the sub target to which the dialogue planner 131 caught failure in a dialogue with a user with the failure at the time of dialogue regulation activation, and failed in it also in such a case and there is an applicable dialogue regulation in it, it will retry.

[0140] In this example, the viewpoint as a human machine interface of voice input has succeeded in \*\*\*\*\* continuing a dialogue with a user by choosing and retrying from things input media called handwriting

recognition as failure occurred in the input media which are inferior in the viewpoint of the soundness of input process, especially the input process of the name of a place which a reading mistake tends to generate.

[0141] Such control is realized as follows. As already stated, the condition of the plan storage section in the event of the system utterance S14, i.e., the utterance "whether it is the schedule which travels what kind of place", shown in gardenia fruit NARIO G of drawing 8 has become like drawing 11.

[0142] Here, suppose that a user is going to nominate "Ryoan-ji" and changed between "\*\*\*\*\*" and reading in the input process in spoken language Mohd. Supposing this "\*\*\*\*\*" does not serve as an object vocabulary of speech recognition, speech recognition processing will fail in it.

[0143] For this reason, a sub target "p4301331" goes wrong and "input process" of that parent target "p430133" is retried under another applicable dialogue regulation "r19." This "r19" is the dialogue regulation of asking for reinput by the same media allocation, and under this dialogue regulation, "p4301331" is developed by the sub target sequence [p43013301, p43013302] as shown in drawing 14, and the dialogue of S15 and U9 is made between equipment and a user in those activation processes.

[0144] that is, equipment: " -- I need your help once again -- " User: "it is \*\*\*\*\*"

The becoming dialogue is made.

[0145] However, speech recognition processing goes wrong for the same reason, and a sub target "p430133" goes wrong again. Next, by making an input mistake a key, the applied regulation "r20" is a dialogue regulation which makes media allocation redo and reinput, and "p430133" is retried by this dialogue regulation and it is developed by the sub target train [p43013311, p43013312, p43013313].

[0146] A sub target "MA selection (input failure (planned sightseeing site)) of p4301311" is further developed by the sub target sequence [p430133111, p430133112] under a dialogue regulation "r12."

[0147] "MA inquiry" of a sub target "p430133111" is a dialogue primitive, and the call of the media assignment-decision section 12 is performed according to the content of the dialogue primitive storage table 133.

[0148] As already stated, with reference to the media [ degree ] assignment preference table 122 showing the media assignment-decision

section 12 in drawing 6 , a media [ degree ] assignment preference node "m6" is chosen in this case. This media [ degree ] assignment preference node is selection of media allocation of changing only the input mode of the present media allocation into pen character string input mode from spoken language Mohd, and by continuing the dialogue of a variable input according to this media allocation, it goes ahead with a dialogue after S16 of drawing 8 , and it becomes possible to return to a dialogue which makes the task of selection of accommodations continue further.

[0149] Namely, equipment: "it has not recognized. Please write the identifier of a destination with a pen."

Character recognition is carried out because a user writes it to a screen as "Ryoan-ji" with a pen, and equipment recognizes "Ryoan-ji" to be "\*\*\*\*\*".

Equipment: "it is Ryoan-ji (\*\*\*\*\*). Since it searches, please wait for a while."

The becoming dialogue is successful.

[0150] According to this equipment, failure is recovered by having considered as the configuration controlled to change an I/O means in this way that it should change to another I/O means and should carry out again to failure in a dialogue with a user, and it becomes possible to continue a dialogue with a user.

[0151] furthermore , in case media allocation will be perform in the media allocation generation section 121 if the input mode components which add the device which catch external noise to this equipment , and judge the felicity of a voice input/output by the input be realize so that it may be cope with , when an external noise become intense , for example , be choose be lose about the media allocation describe that voice input be possible or that a voice output be possible by the constraint section of the Mohd-media conversion table of drawing 7 .

[0152] Consequently, while the media allocation which performs voice input is chosen and the input from a user is received with voice, when the noise becomes intense suddenly, [ for example, ] Although a media assignment-decision section call occurs with speech recognition failure like the above-mentioned example of operation and selection of media [ degree ] allocation in the media assignment-decision section is newly carried out, being chosen of media allocation which uses voice for I/O in the selection is lost.

[0153] Since it is used in this way by having added the control which changes an I/O method according to change of an external-

environment-element according to an external-environment-element according to the equipment of such a configuration, choosing the optimal I/O method, it becomes possible to continue a dialogue with a user.

[0154] In addition, in an above-mentioned example, it has three modules, the input device processing section 111, the input mode processing section 112, and the input media processing section 113, about the configuration of the input section 11. Although it has described having three modules, the output media processing section 141, the output mode processing section 142, and the output device processing section 143, also about the output section 14 as a premise It is not limited to this and the configuration of the radial transfer components which realize this equipment can perform various deformation.

[0155] Moreover, although the above-mentioned example has described as another component with which the input section and the output section became independent, realizing as the I/O section which unified these is also possible, and such an implementation approach is also within the limits of the meaning of this invention.

[0156] Moreover, although the example mentioned above has described the media assignment-decision section as the component part which became independent of the dialogue Management Department or a problem-solving means, and components which became independent to the input section, the output section, the dialogue Management Department, etc. similarly about the problem-solving means The configuration approach of components of realizing this equipment is not what is limited to this. For example, it is possible to realize this equipment, even if there is no media assignment-decision section as components which became independent by realizing the dialogue Management Department with a media assignment-decision function. Moreover, for example, a problem-solving means is possible also for the function of the dialogue Management Department, the function of the media assignment-decision section, and constituting so that it may have the function of the input section, and the function of the output section further, and such an implementation approach is also within the limits of the meaning of this invention.

[0157] Thus, the equipment which realizes this invention has the various implementation approaches about the bill of materials, and it refuses that it is not what is limited to the configuration of the components explained here. Moreover, although the example mentioned above has described as a premise using the media [ degree ] assignment preference table 122 which determines the next media allocation as the

media allocation at a certain event based on the situation in the event about the control system of the media assignment-decision section 12. The media assignment-decision method which realizes this equipment is not limited to this. For example, it also becomes possible to realize selection of finer media allocation by being able to realize also in the media assignment-decision method based on a state-transition model, and adding slots and information, such as whenever

[ preference ], to the media [ degree ] assignment preference table 122.

[0158] Moreover, although one batch, i.e., current media allocation information, is clearly described by the example mentioned above as history information of the media allocation used in case it opts for media allocation in the media allocation generation section 121, it is possible by describing the condition of the past media allocation, for example to the condition part of the media [ degree ] assignment preference table 122 to use the history information of media allocation to the decision of media allocation.

[0159] Moreover, in an above-mentioned example, although the dialogue management based on a hierarchy plan generation method is described as a premise, the dialogue management method which realizes this equipment is not limited to this, and can be realized also in the dialogue control system based on a state-transition model.

[0160] Furthermore, although the user is making the user know by displaying the available input approach with other information on a display in an above-mentioned example, the presentation format of the input media which realize this equipment is not limited to this, and can be realized also as a thing to show from other output devices, such as displaying on another display for example.

[0161] Moreover, although the user is making the user learn only the available input approach by displaying on a display in an above-mentioned example Even if this equipment performs the response with voice to a user hard of hearing possible [ that making a user know is not limited to the input approach, and it displays similarly about an output method ] Since a user can know it, a user becomes possible [ requiring modification of media allocation ].

[0162] Furthermore, for example, the specific user model storage section is prepared in the media assignment-decision section, and it becomes possible to set up the optimal media allocation corresponding to each user, without a user requiring each time, when it receives a user hard of hearing [ above-mentioned ] because a user records beforehand available media or the media which cannot be used there.

[0163] Moreover, for example, record the situation of user each about the above media allocations on an adjustable medium like an IC card, for example, this equipment is equipped with the device in which IC card read is possible. By a user making that IC card insert and read using this device at the time of utilization initiation, and making the specific user model storage section memorize that read information Also in the equipment for many and unspecified persons, the situation of user each can be set up beforehand, and it can be coped with now, consequently the equipment which becomes possible from the event of setting out of the media allocation according to the situation of user each being the beginning of utilization initiation comes to be obtained. In this case, without carrying out demand actuation of modification, since it carries out from the event of each user's optimal media allocation being the beginning of utilization initiation, since smooth utilization is attained, it is useful to the improvement in service.

[0164] Moreover, in being SHISUTEKE which only a specific user uses, it is made to perform specific user model storage to nonvolatile memory, and a user can enable it to update, or it can take for repeating utilization by the thing of giving a learning function and making it make a study result reflect the content of storage etc., and the interface which the user tends to use can be provided now.

[0165] The interface which the user tends to use gradually can be offered only by a user not updating a user model by himself, but repeating utilization also for \*\* by equipping the media assignment-decision section with functional learning functions, such as ID3 and neuro, furthermore, and carrying out renewal of automatic of the content of the specific user model storage section by study.

[0166] Moreover, although the specific user considered as available media or the media which cannot be used about the content of storage of the specific user model storage section above The content of storage of the specific user model storage section is not what is limited to this. For example, an interface desirable for each user can be similarly offered now by memorizing in the specific user model storage section about the media which a specific user selects, and such an implementation approach is also within the limits of the meaning of this invention.

[0167] It is also possible to adopt the approach of making a user know with determining the output media for making the approach which is still more nearly available learn itself in the media assignment-decision section, and outputting using the output media allocation based on the

decision, and such an implementation approach is also within the limits of the meaning of this invention.

[0168] In addition, although the above-mentioned example showed the example treating failure in the speech recognition by a user's reading mistake for example, like [ when an environmental sound (noise) becomes a cause and voice input is checked ] Also in a situation which injury generates in a dialogue by change of the situation of the media to be used The media monitoring function which supervises the condition of an input and output media is included in the media change section. It is set up by the media change section at a certain event, and the media allocation which is going to carry out current utilization is used without a failure, and it becomes possible by supervising whether communication with a user and dialogue equipment was attained safely to perform processing for clear trouble.

[0169] Namely, for example, it sets in the scene where it receives the voice input from a user, the scene where voice tends to perform the output to a user, etc. When inhibition of communication with the user and dialogue equipment by generating of the loud noise etc. arises by monitoring perimeter noise level etc. continuously, for example, the dialogue Management Department is notified of this. For example, for example In a certain failure generating, stopping the I/O with voice and making it make a screen display and keyboard entry perform etc. makes a media change appropriately, or or -- countering -- etc. -- the dialogue equipment which it becomes possible to realize processing of starting the sub dialogue for clear trouble, consequently can cope with it also to an environmental change is realizable. Thus, in this invention, it is possible to perform deformation various in the range which does not deviate from the meaning.

[0170]

[Effect of the Invention] As explained above, according to this invention, it becomes possible by having chosen suitable I/O media dynamically according to the available I/O means at the situation of a dialogue, and its event to provide a user with the interface which is easy to use according to a situation. Moreover, it can know easily with what kind of means a user can input, and it becomes possible to offer the interface which is easier to use.

[0171] moreover, the quality and the amount of the information which application demands in an output -- corresponding -- a user -- accepting -- or [ \*\*\*\*\* ] -- being easy -- it becomes possible to set up the combination of an output means, it becomes possible to offer the

interface which is easy using for the user of each in using the I/O approach which a user demands, and it becomes that it is possible to offer the interface which is easy using for each user by using a desirable I/O means for a user. Moreover, it becomes possible to offer the interface which is easy to use for each user by avoiding the I/O means which is not desirable for a user. Furthermore, failure in a dialogue with a user is recovered by changing an I/O means, and practically great effectiveness, like it becomes possible to continue a dialogue with a user is done so.

---

[Translation done.]

\* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
  2. \*\*\*\* shows the word which can not be translated.
  3. In the drawings, any words are not translated.
- 

## DESCRIPTION OF DRAWINGS

---

### [Brief Description of the Drawings]

[Drawing 1] The block block diagram for being drawing for explaining this invention and explaining the mode of concrete operation of this invention.

[Drawing 2] Drawing in which being drawing for explaining this invention and showing the example of the dialogue regulation for explaining the mode of concrete operation of this invention.

[Drawing 3] Drawing in which being drawing for explaining this invention and showing the example of a dialogue primitive storage table for explaining the mode of concrete operation of this invention.

[Drawing 4] Drawing in which being drawing for explaining this invention and showing the example of the plan storage section (at the time of starting) for explaining the mode of concrete operation of this invention.

[Drawing 5] Drawing in which being drawing for explaining this invention and showing the example of the plan storage section (at the time of standby) for explaining the mode of concrete operation of this invention.

[Drawing 6] Drawing in which being drawing for explaining this invention and showing the example of the media [ degree ] assignment preference table for explaining the mode of concrete operation of this invention.

[Drawing 7] Drawing in which being drawing for explaining this invention and showing the example of the Mohd-media allocation conversion table for explaining the mode of concrete operation of this invention.

[Drawing 8] It is drawing about the example of the example scenario of operation which is drawing for explaining this invention and is assumed for explaining the mode of concrete operation of this invention.

[Drawing 9] Drawing in which being drawing for explaining this invention and showing the example of the plan storage section (at the time of an initial condition input request) for explaining the mode of concrete operation of this invention.

[Drawing 10] Drawing in which being drawing for explaining this invention and showing \*\*\*\* of the plan storage section (at the time of initial retrieval termination extract) for explaining the mode of concrete operation of this invention.

[Drawing 11] Drawing in which being drawing for explaining this invention and showing the example of the plan storage section (a condition narrowing-down dialogue, extract) for explaining the mode of concrete operation of this invention.

[Drawing 12] Drawing in which being drawing for explaining this invention and showing the example of the plan storage section (at the time of a user interrupt extract) for explaining the mode of concrete operation of this invention.

[Drawing 13] Drawing in which being drawing for explaining this invention and showing the example of the plan storage section (at the time of user interrupt management extract) for explaining the mode of concrete operation of this invention.

[Drawing 14] Drawing in which being drawing for explaining this invention and showing the example of the plan storage section (dealing with speech recognition failure, extract) for explaining the mode of concrete operation of this invention.

[Drawing 15] The detailed block diagram for being drawing for explaining this invention and explaining the mode of concrete operation of this invention.

[Drawing 16] Drawing in which being drawing for explaining this invention and showing the example of a screen of the media allocation "Hugh Mann" for explaining the mode of concrete operation of this invention.

[Drawing 17] Drawing in which being drawing for explaining this invention

and showing the example of a screen of the media allocation "a variable display voice input/output" for explaining the mode of concrete operation of this invention.

[Drawing 18] Drawing in which being drawing for explaining this invention and showing the example of a screen of the media allocation "touch panel display selection and element name voice reading raising" for explaining the mode of concrete operation of this invention.

[Drawing 19] Drawing in which being drawing for explaining this invention and showing the example of an algorithm of the dialogue management for explaining the mode of concrete operation of this invention.

[Drawing 20] Drawing in which being drawing for explaining this invention and showing the example of a standby condition for explaining the mode of concrete operation of this invention.

[Drawing 21] Drawing in which being drawing for explaining this invention and showing the example of a screen of the media allocation "tabular format touch panel display selection / retrieval result number-of-cases voice output" for explaining the mode of concrete operation of this invention.

#### [Description of Notations]

10 -- Multi-modal dialogue equipment

11 -- Input section

12 -- Media assignment-decision section

13 -- Dialogue Management Department

14 -- Output section

111 -- Input device processing section

112 -- Input mode processing section

113 -- Input media processing section

121 -- Media allocation generation section

122 -- Media [ degree ] assignment preference table

123 -- The present media allocation storage slot

124 -- Mode media allocation conversion table

131 -- Dialogue planner

132 -- Dialogue regulation table

133 -- Dialogue primitive storage table

141 -- Output media processing section

142 -- Output mode processing section

143 -- Output device processing section

143a -- Output media distribution switch

143b -- Output media processing section

143c -- The Media Mohd interface

SW -- Media selecting switch  
 IDVa-IDVn -- Input device  
 ODVa-ODVn -- Output device

[Translation done.]

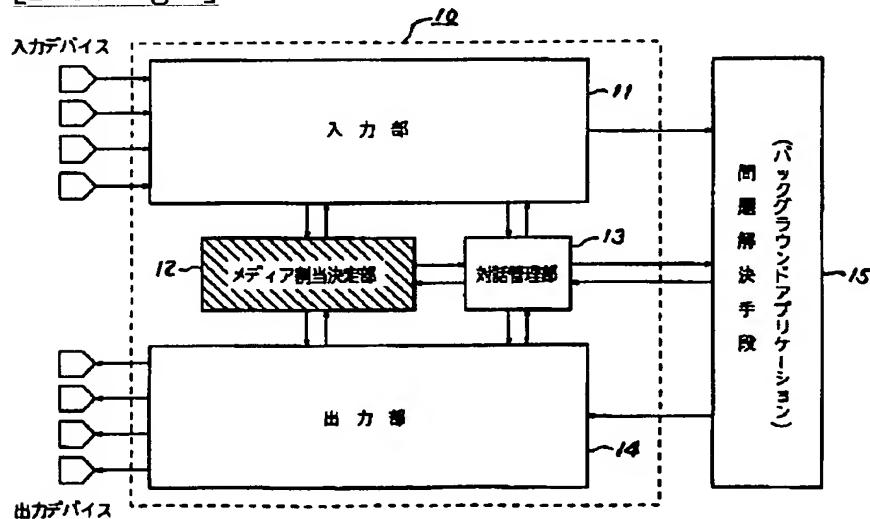
\* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

DRAWINGS

[Drawing 1]



[Drawing 3]

ファクトID	目標	仕様
f01	MA固合せ(条件)	メディアアロケータ呼びだし(条件)
f02	検出特機	
f03	MA設定(MA)	
f04	変数入力(変数)	入力メディア処理部呼び出し(入力要求(変数))

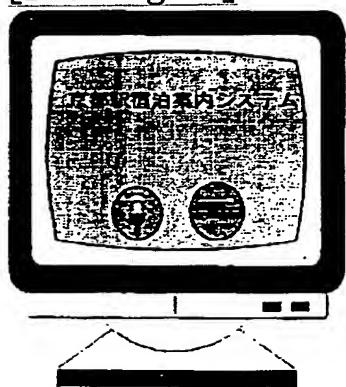
対話プリミティブ記憶テーブル

[Drawing 16]



メディア割当「ヒュ・マン」の画面例

[Drawing 20]



待機状態

[Drawing 2]

ルールID	目標	摘要条件	サブ目標リスト
r00	対話セッション		ユーザ検出, MA選択(初期), 開始セッション, タスク実行, 終了セッション
r01	タスク実行		初期条件入力、検索、候補決定、候補確認
r02	初期条件入力		入力要求 (A→D, 入力変数リスト)、MA選択 (入力要求 (入力変数リスト))、入力処理 (入力変数リスト)。
r03	開始セッション		開始要求 (U→S), 開始確認 (S→U)
r04	開始セッション		開始要求 (S→U), 開始確認 (U→S)
r05	終了セッション		終了要求 (U→S), 終了確認 (S→U)
r06	終了セッション		終了確認 (S→U)
r07	候補決定	候補数 = 1	
r08	候補決定		選択要求
r09	候補決定		検索条件取り込み、検索、候補決定。
r10	選択要求		MA選択、選択肢表示、ユーザ選択。
r11	ユーザ検出		MA選択(待機)、検出待機。
r12	MA選択(条件)		MA選択(条件, MA1), MA選択(MA1)。
r13	表現形式変更要求 (U→S)		MA選択、データ出力。
r14	検索条件取り込み		未入力変数組合せ (D→A, 未入力変数リスト)、number (入力変数、入力変数リスト)、 追加条件入力 (入力変数、[検索込み])。
r15	選択要求	MA要求 (MA, U→S)	MA選択 (ユーザ要求 (MA)), 選択表示、ユーザ選択。
r16	選択要求	検索条件要求 (条件)	MA選択 ( )。
r17	追加条件入力 (入力変数、入力目的)		MA選択 (追加条件入力 (入力変数、入力目的))、 入力目的表示 (D→U, [追加条件入力   入力目的])、 入力処理 (入力変数)。
r18	入力処理 (入力変数)		変数入力 (入力変数)。
r19	入力処理 (入力変数)		再入力要求。 变数入力 (入力変数)。
r20	入力処理 (入力変数) 再入力要求。		MA選択 (入力失敗 (入力変数))。 变数入力 (入力変数)。

### 対話規則

#### [Drawing 4]

ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01, p02, p03, p04, p05]	
p01	ユーザ検出	処理中	[ ]	
p02	MA選択(初期)	未処理		
p03	開始セッション	未処理		
p04	タスク実行	未処理		
p05	終了セッション	未処理		

### プラン記憶部 (起動時)

## [Drawing 5]

ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01,p02,p03,p04,p05]	
p01	ユーザ検出	処理中	[p11,p12]	
p02	MA選択(初期)	未処理		
p03	開始セッション	未処理		
p04	タスク実行	未処理		
p05	終了セッション	未処理		
p11	MA選択(待機)	成功	[p111,p112]	待機
p12	検出待機	処理中	[ ]	
p111	MA開始(待機,MA1)	成功	[ ]	MA1=待機
p112	MA設定(待機)	成功	[ ]	

プラン記憶部(待機時)

## [Drawing 6]

ノードID	現メディア割当	条件	次メディア割当
m01	---	呼出条件=待機	待機
m02	---	呼出条件=初期	ヒューマン
m03	---	呼出条件=入力要求	实效音声入力画面音声出力
m04	---	呼出条件=選択要求 & 要素数<=4	タッチパネル表示選択・要素名音声読み上げ
m05	---	呼出条件=選択要求 & 5<=要素数<=100	場形式タッチパネル表示選択・複数結果件数音声出力
m06	(音声言語, XXX)	呼出条件=入力失敗	(ペン文字列, XXX)

次メディア割当選択テーブル

[Drawing 7]

ノードID	メディア割当名	機能	入力モード組
mm01	ヒューマン	音声入出力可能	( 音声電話・音声電話+顔画像 )
mm02	待機	入力センサ利用可能	( センサ検出特機・待機メッセージ )
mm03	変数音声入力画面音声出力	変数型 ∈ (DATE, INT, MONEY) & 入力変数数 <= 4 & 音声入力可能	( 音声電話・音声電話+変数画面表示 )
mm04	ペン文字列入力・ヒューマン 対応	音声出力可能	( ペン文字列・音声電話+顔画像 )
mm05	タッチパネル選択・要素名 音声読み上げ	要素数 = N	( N要素タッチパネル(並列)・音声電話 ( N件 ) + N要素並列配置 )
mm06	投影式タッチパネル表示選択 ・検索結果件数音声出力	要素数 = N	( N要素タッチパネル(表示式)・音声電話 ( N件 ) + N要素表示 )

モード・ディア割当対応表

[Drawing 8]

A)	案内システムの前に一人の旅行者が立つ センサが反応してディスプレイがオンになり、システムのタイトル画面 効果音とともに入力を促すプロンプトが出る。
B)	旅行者が音声を使ってシステムに働きかける 利用者 (U2): 「宿を探してください」 システム (S1): 「了解しました」
C)	初期データ獲得対話 (S2): 「いつからご宿泊ですか？」 (U2): 「今日からお願いします」 (S3): 「ご出立はいつになりますか？」 (U3): 「あさっての予定です」 (S4): 「何部屋必要ですか？」 (U4): 「一人ですので一部屋で結構です」 (S5): 「一泊あたりのご予算はいくらぐらいですか？」 [画面に示す選択肢の中から選んで下さい] (U5): 「6000円以下でお願いします」 (U6): 「一泊6000円程度までの宿で本日から12日水曜日までの3泊。 一部屋のご予約を御希望ですね。」 (U7): 「では、ただ今から検索しますのでしばらくお待ち下さい。」
D)	候補の表示 (S8): 「お待たせいたしました」 (S9): 「おっしゃった条件で本日から宿泊可能な宿は東北の 雄山山荘、ひなびた庵、奥山院の3件です。」 (S10): 「各宿泊施設の詳しい情報を御観になりたい場合は、その名前をしゃべるか、 画面の該当部分を指で押して下さい。」
E)	利用者主導の対話 (U6): 「それぞれ道順と時間を教えて下さい」 (S11): 「雄山山荘はバスで約1時間半ほどです。 ひなびた庵には列車で30分ほど行ったところから バスに乗り換えて30分ほどかかります。 奥山院はここからバスで30分ほど乗ってから 30分ほど歩いたところにあります」
G)	検索挿り込みの提案 (S13): 「件数がS34とまだ多いので、条件を挿り込みたいと思います」 (S14): 「どのようなところを観光される御予定でしょうか？」
H)	ユーザの読み間違い(音声認識失敗)への対処 (U8): 「りゅうあんじです」 (S15): 「は？もう一度お願いします」 (U9): 「りゅうあんじです」 (S16): 「認識できませんでした。ベンで行き先の名前を書いてください」 (S17): 「竜安寺(りょうあんじ)ですね。検索しますのでしばらくお待ち下さい」

## 想定する動作別シナリオ

[Drawing 9]

ノードID	名稱	効果	サブノードIDリスト	結果
p0	对话セッション	処理中	[p01, p02, p03, p04, p05]	
p01	ユーザ検出	成功	[p11, p12]	
p02	MA認証(初期)	成功	[p11, p12]	
p03	開始セッション	成功	[p31, p32]	
p04	タスク実行	成功	[p41, p42, p43, p44]	待機
p05	終了セッション	未処理		
p11	MA選択(待機)	成功	[p111, p112]	MA1=待機
p12	検出待機	成功	[ ]	
p111	MA融合(待機, MA1)	成功	[ ]	MA2=ヒューマン
p112	MA選択(待機)	成功	[ ]	
p21	MA融合(待機, MA2)	成功	[ ]	
p22	MA選択(ヒューマン)	成功	[ ]	
p31	開始要求(U→S)	成功	[ ]	
p32	開始確認(S→U)	成功	[ ]	
p41	初期条件入力	未処理	[p411, p412, p413]	
p42	状況決定	未処理		
p43	候補確認	未処理		
p44	候補確認	未処理		
p411	入力要求(A→D, [端日, 出発日, 人量, 予算])	成功	[ ]	
p412	MA選択(入力要求([端日, 出発日, 人量, 予算]), MA3)	成功	[p4111, p4112]	
p412	入力確認([端日, 出発日, 人量, 予算])	未処理		
p4111	MA融合(入力要求([端日, 出発日, 人量, 予算]), MA3)	成功	[ ]	
p4112	MA選択(入力要求([端日, 出発日, 人量, 予算]))	成功	[ ]	MA1=最初の待機入力

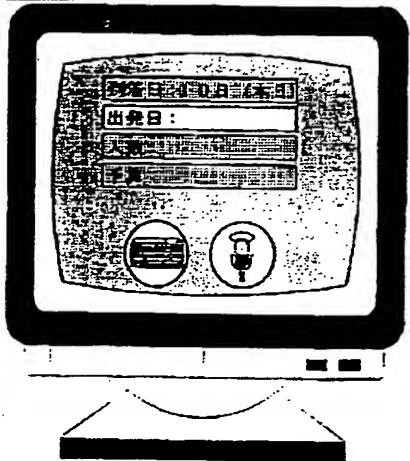
プラン記憶部（初期条件入力要求待）

[Drawing 10]

ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01,p02,p03,p04,p05]	
p01	ユーザ検出	成功	[p11,p12]	
p02	MA選択(初期)	成功	[p21,p22]	
p03	開始セッション	成功	[p31,p32]	
p04	タスク実行	処理中	[p41,p42,p43,p44]	
p05	終了セッション	未処理		
...中略...				
p41	初期条件入力	成功	[p411,p412,p413]	
p42	検索	成功	[ ]	
p43	候補決定	処理中	[p431]	
p44	候補確認	未処理		
...中略...				
p431	選択要求	処理中	[p4311,p4312,p4313]	
p4311	MA選択(選択指示)	処理中	[p43111,p43112]	
p4312	タスク実行	未処理		
p4313	終了セッション	未処理		
p43111	MA選択(初期)	処理中	[ ]	
p43112	開始セッション	未処理		

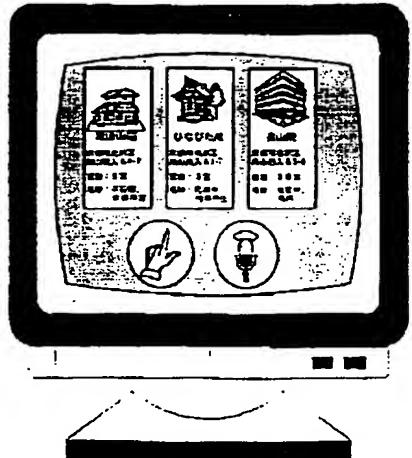
### プラン記憶部(初期検索終了時、抜粋)

#### [Drawing 17]



メディア割当'最初未音声入出力'の画面例

#### [Drawing 18]



ノーティア・クッティパネル表示選択・要素名音声読み上げの画面例

[Drawing 11]

ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01,p02,p03,p04,p05]	
p01	ユーザ検出	成功	[p11,p12]	
p02	MA選択(初期)	成功	[p21,p22]	
p03	開始セッション	成功	[p31,p32]	
p04	タスク実行	処理中	[p41,p42,p43,p44]	
p05	終了セッション	未処理		
-中略-				
p41	初期条件入力	成功	[p411,p412,p413]	
p42	検索	成功	[]	
p43	候補決定	失敗	[p431]	
p44	候補確認	未処理		
-中略-				
p431	選択要求	失敗	[p4311,p4312,p4313]	
p4311	MA選択(選択肢提示)	失敗	[p43111,p43112]	
p4312	選択肢提示	未処理		
p4313	ユーザ選択	未処理		
p43111	MA選択(選択肢提示, MA 1)	失敗	[]	
p43112	MA設定(MA 1)	未処理		
p43	候補決定	再試行	[p4301,p4302,p4303]	
p4301	検索条件挿り込み	処理中	[p43011,p43012,p43013]	
p4302	検索	未処理		
p4303	候補決定	未処理		
p43011	未入力変数埋め( D→A, 未入力変数リスト )	成功	[]	未入力変数リスト=[観光地名, 領域名, 駐車場]
p43012	member( 入力変数, [ 観光予定地, 領域名, 駐車場 ] )	成功		入力変数リスト
p43013	追加条件入力( 観光予定地, [ 検索挿り込み ] )	処理中	[p4301311,p4301312]	
p430131	MA選択( 追加条件入力( 観光予定地, [ 検索挿り込み ] ) )	成功	[p4301311,p4301312]	
p430132	入力目的提示( D→U, [ 追加条件入力, 検索挿り込み ] )	成功		
p430133	入力処理( 観光予定地 )	処理中	[p4301331]	
p4301331	変数入力( 観光予定地 )	処理中	[]	
-中略-				

### プラン記憶部（条件較り込み対話・操作）

### [Drawing 12]

ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01,p02,p03,p04,p05]	
p01	ユーザ検出	成功	[p11,p12]	
p02	MA選択(初期)	成功	[p21,p22]	
p03	開始セッション	成功	[p31,p32]	
p04	タスク実行	処理中	[p41,p42,p43,p44]	
p05	終了セッション	未処理		
...中略...				
p41	初期条件入力	成功	[p411,p412,p413]	
p42	検索	成功	[]	
p43	候補決定	処理中	[p431]	
p44	候補確認	未処理		
...中略...				
p431	選択要求	処理中	[p4311,p4312,p4313]	
p4311	MA選択(選択提示)	成功	[p43111,p43112]	
p4312	選択肢提示	成功	[]	
p4313	ユーザ選択	未処理	[]	
...中略...				

### プラン記憶部(利用者登録時抜粋)

#### [Drawing 21]



メディア端末「複数式タッチパネル表示選択 検索結果件数音声エラー」の品番例

#### [Drawing 13]

ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01,p02,p03,p04,p05]	
p01	ユーザー検出	成功	[p11,p12]	
p02	MA選択(初期)	成功	[p21,p22]	
p03	属性セッション	成功	[p31,p32]	
p04	タスク実行	処理中	[p41,p42,p43,p44]	
p05	終了セッション	未処理		
***中略***				
p11	初期条件入力	成功	[p411,p412,p413]	
p12	検索	成功	[ ]	
p43	候補決定	処理中	[p431]	
p44	候補選択	未処理		
***中略***				
p431	選択要求	失敗	[p4311,p4312,p4313]	
p4311	MA選択(選択提示)	失敗	[p43111,p43112]	
p4312	選択表示	失敗	[ ]	
p4313	ユーザー選択	失敗	[ ]	
***中略***				
p431	選択要求	取扱行	[p43101,p43102,p43103]	
p43101	MA選択(ユーザー要求(出力([選択, 到達時間]))	成功	[p431011,p431012]	
p43102	選択表示	未処理	[ ]	
p43103	ユーザー選択	未処理		
p431011	MA候補(ユーザー要求(出力([選択, 到達時間])))	成功	[ ]	
p431012	MA選択(MA1)	成功		
	MA選択(MA1)	成功		

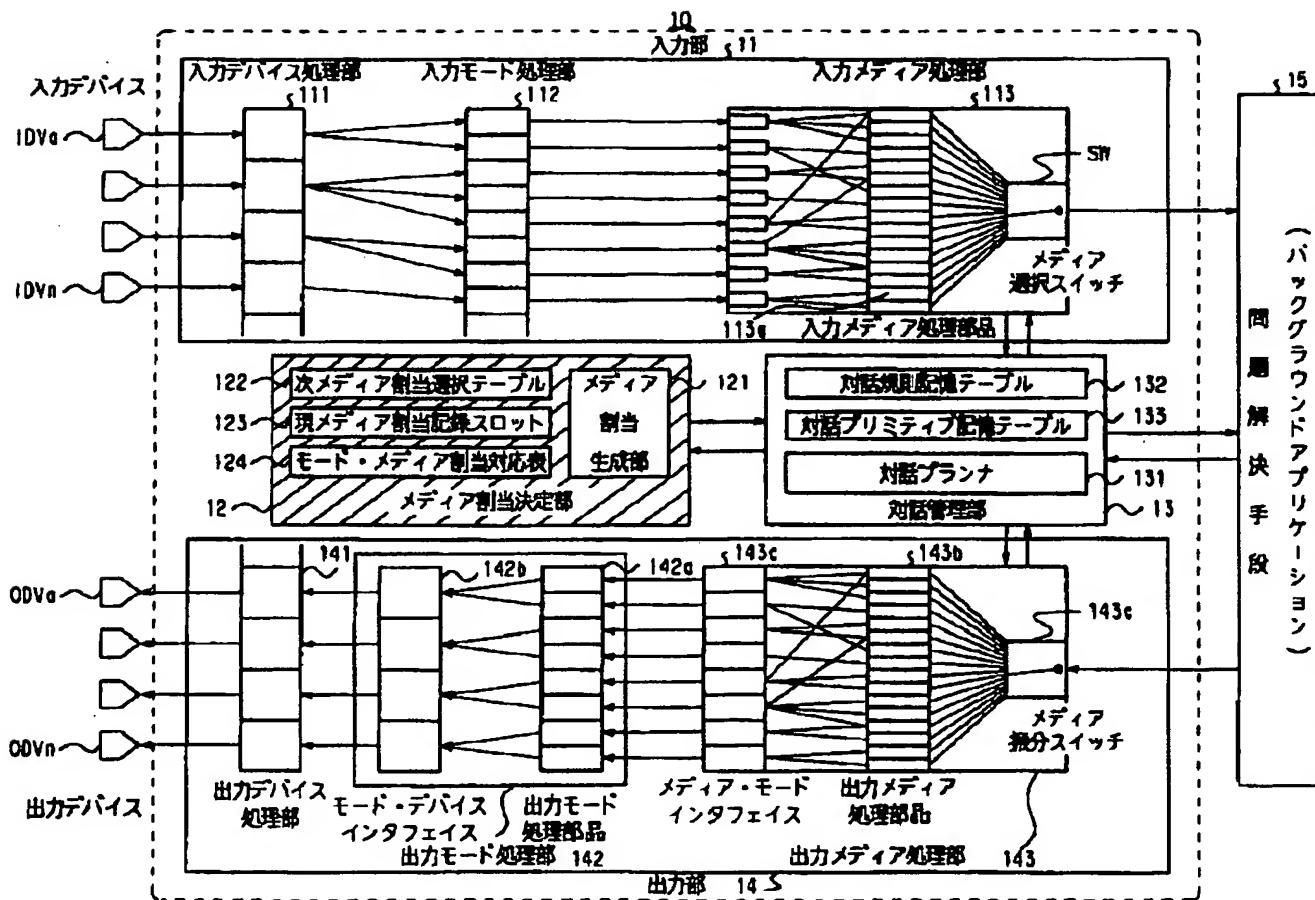
プラン記憶部(利用者登録登録時、接続)

[Drawing 14]

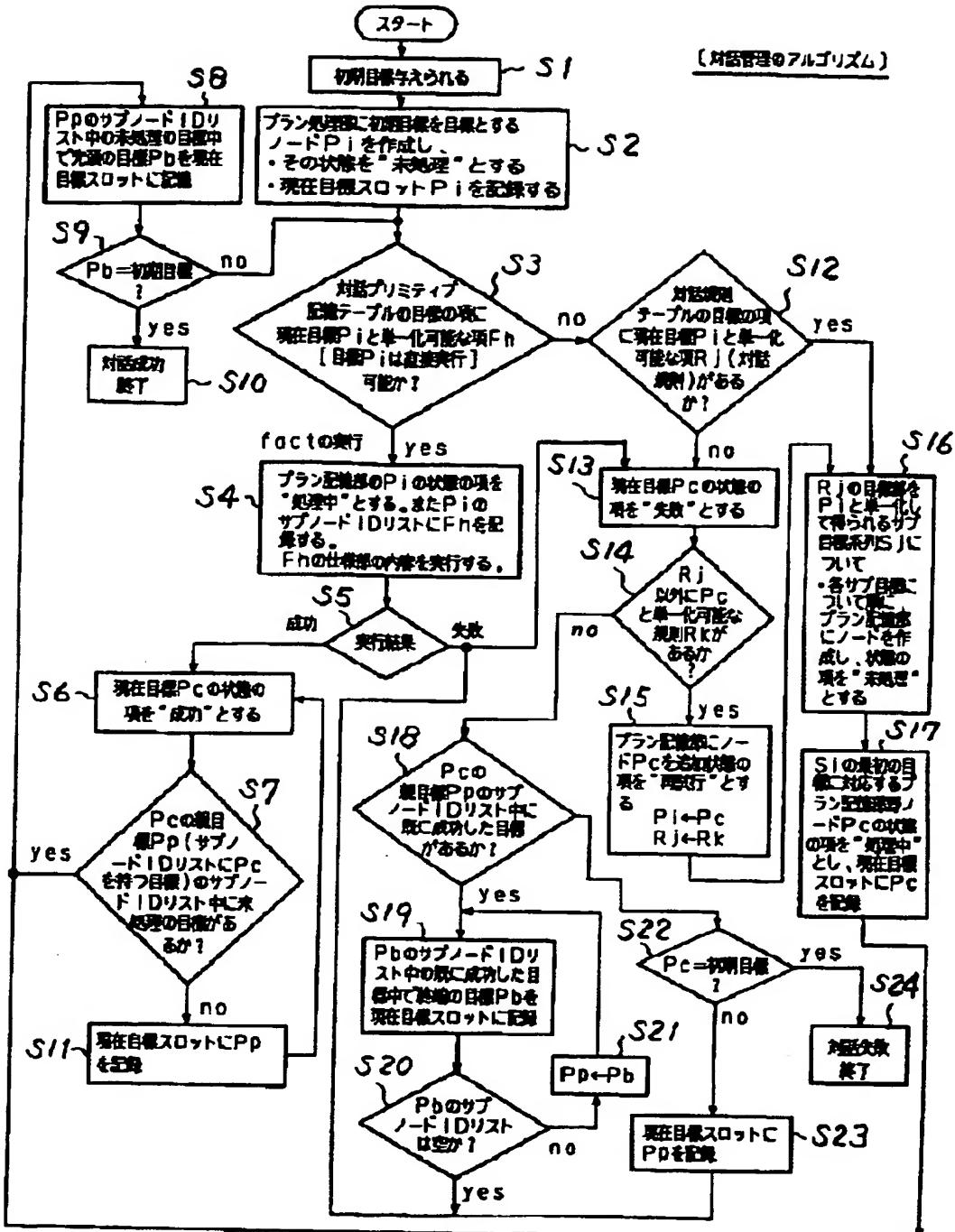
ノードID	目標	状態	サブノードIDリスト	結果
p0	対話セッション	処理中	[p01,p02,p03,p04,p05]	
p01	ユーチューブ検出	成功	[p11,p12]	
p02	MA選択(初期)	成功	[p21,p22]	
p03	開始セッション	成功	[p31,p32]	
p04	タスク実行	処理中	[p41,p42,p43,p44]	
p05	終了セッション	未処理		
---中略---				
p41	初期条件入力	成功	[p411,p412,p413]	
p42	検索	成功	[]	
p43	候補決定	失敗	[p431]	
p44	候補部	未処理		
---中略---				
p431	選択要求	失敗	[p4311,p4312,p4313]	
p4311	MA選択(選択肢提示)	失敗	[p43111,p43112]	
p4312	選択肢提示	未処理		
p4313	ユーチューブ選択	未処理		
p43111	MA結合(選択肢提示,MA1)	失敗	[]	
p43112	MA設定(MA1)	未処理		
p43	候補決定	再実行	[p4301,p4302,p4303]	
p4301	検索条件取り込み	処理中	[p43011,p43012,p43013]	
p4302	検索	未処理		
p4303	候補決定			
p43011	未入力候補合せ(D→A,未入力変数リスト)	成功	[]	未入力候補合せ(D→A,未入力変数リスト)
p43012	member(入力変数, [観光予定地,選択肢式,複数選択肢式])	成功		
p43013	選択条件入力(観光予定地, [検索取り込み])	処理中	[p4301311,p4301312]	
p430131	MA選択(追加条件入力(観光予定地, [検索取り込み]))	成功	[p4301311,p4301312]	
p430132	入力目的提示(D→U, [追加条件入力, 検索取り込み])	成功		
p430133	入力処理(観光予定地)	失敗	[p4301331]	
p4301331	変数入力(観光予定地)	失敗	[]	
---中略---				
p430133	入力処理	失敗	[p43013301,p43013302]	
p43013301	最入力要求	失敗	[]	
p43013302	変数入力(観光予定地)	失敗	[]	
p430133	入力処理(観光予定地)	再実行	[p43013311,p43013312,p43013313]	
p43013311	MA選択(入力失敗(観光予定地))	処理中	[p430133111,p430133112]	
p43013312	再入力要求	未処理		
p43013313	変数入力(観光予定地)	未処理		
p430133111	MA結合(入力失敗(観光予定地),MA1)	処理中	[]	
p430133113	MA設定(MA1)	未処理		

プラン記憶部(音声記憶失敗への対処, 技术)

[Drawing 15]



### [Drawing 19]



[Translation done.]